MapPapers 1en-II, 2012, pp.21-38
doi:10.4456/MAPPA.2012.02

A database for archaeological data recording and analysis

Anichini F., Fabiani F., Gattiglia G., Gualandi M.L.

The archaeological database is managed within an RDBMS aimed at describing the historical, stratigraphic, urban and environmental complexity of the urban area of Pisa, and directly connected to a digital archive containing the raw data. The RDBMS structure is based on four different logical levels that gradually manage the information through an interpretative synthesis process: from definition of the material trace to classification of data into typological and chronologically-divided macro categories which have archiving and analysis functions, as well as categorisation purposes for the predictive calculation of archaeological potential.

Keywords: RDBMS; excavation archaeology; archaeological interventions; heterogeneity; categorisation.

The digital archive used for recoding archaeological excavation data is part of a larger archive which, alongside archaeological data, also includes data from aerial archaeology, from building archaeology, and from the analysis of written documentation and toponymy. The archive is composed of an RDBMS and a series of directories specifically structured on the laboratory server file[1]. The archive of archaeological excavation data also has the double role of archiving and analysing the data[2] and of categorising them for the predictive calculation of archaeological potential (Bini D., Dubbini N., Steffè S. 2011). The directories containing all archived files are symmetrical with respect to the RDBMS[3].

The part of the RDBMS that deals with the recording and analysis of archaeological excavation data held by the MAPPA archaeological research team will be described in this article. Although this is an internal work tool and public disclosure is not expected, we believe it is important to provide information and share our considerations and the solutions adopted. Our aim, therefore, is not to analyse the methodological and epistemological problems inherent to the design of an archaeological database[4], nor to give a detailed description of the technical and IT steps that are at the basis of its implementation. Our aim is rather to shed light on the basic guidelines taken when creating the data bank to allow appropriate understanding of the work carried out. Designing an archaeological database is foremost a methodological work. It is the archaeologist's duty, not that of the IT expert, to decide – on the basis of the research project aims – which kind of information and processing must be performed, as well as how this information will be handled by the processor: the type of data, the registration methods and the level of detail and accuracy chosen[5].

The database merges into a RDBMS[6] implemented upon the following principles:

- Creating a container that allows the greatest possible amount of information with partially incon-

1 The overall structure of the archive and its components will be the subject of closer examination.

2 The table-based structure, therefore, can either 'join' or 'relate' to the vector files contained in the GIS.

3 In the case of excavation archaeology, the structure revolves around the INTERVENTIONS directory which contains (when available) excavation reports, context/phase plans, Harris matrix and/or Context records, in different origin or acquisition formats, related to each archaeological excavation. All these files are managed using XnView software which tags single files and folders and allows easier management of the files (the INTERVENTIONS Directory currently contains 33,490 files, divided into 4,292 folders, equal to actual disk space of 80 GB). Indexing of the files required 120 man days (2 months of work shared among 3 operators). Great efforts were required for this work which was however necessary to allow the rapid retrieval of data once registered.

4 Please refer to Fronza 2009a, D'andrea 2006: 48 ss, Gabucci 2005: 30 ss.

5 Fronza 2004: 400

6 For a general overview see Fronza 2003

sistent features to be registered;

- Containing the greatest possible amount of information so that the consultation and analysis activities may take into account all the data collected and produced from the search;
- Being provided with an open architecture which can be easily integrated both in terms of data implementation and structure modification, if necessary;
- Allowing different use according to the user's computer literacy through implementation of a user-friendly interface.

1. Managing heterogeneity

The greatest difficulty encountered during this project phase was synthesising the archaeological data gradually acquired, and making it consistent.

Following evaluation of the different typological cases, three main problems arose:

Heterogeneity of the source of information;

Dissimilarity between the chronological parameters adopted;

Heterogeneity of terminology in the definition of classes and type of finds.

The heterogeneity of sources and of a language (that has moved from "picturesque" to scientific over five centuries of archaeological recording in Pisa), necessarily led to a work of lexical categorisation. In some cases (the implementation of the database is still under way), the work required a strenuous interpretation of the archaeological culture of certain periods[7], as well as the need to redefine outdated chronological categories according to modern parameters.

In order to integrate all the data and avoid the drawbacks due to the strong inconsistency of the above elements, we decided to follow a line of approach that, although arbitrary, took into account the indications provided in the "*Linee guida per la redazione della Carta Archeologica della Toscana* (Guidelines for drawing up the Archaeological Map of Tuscany)" (Francovich, Pellicanò, Pasquinucci 2001: 182-198) and the solutions already tested and adopted in research work on Pisa[8] (Anichini, 2004-2005; Gattiglia 2010; Gattiglia 2011), thus developing an archiving structure based on the urban and peri-urban territorial context subject of our research.

We decided to handle sources equally: source information was synthesised with the minimum unit attributable to the ***archaeological intervention***[9] of any nature whatsoever and with different in-depth relational levels which allowed us to reach (where documented) single contexts and quantification of finds, moving from interpreted data to raw data.

The problem regarding the inconsistency between the chronological parameters adopted over more than a century of urban archaeology in Pisa was addressed on the one hand, by using a chronological classification with the widest diachronic coverage as possible (from pre-history to contemporary ages, which also includes the present day) and, on the other hand, by using accurate chronological parameters. The latter were defined on the basis of *thesauri*[10] managed both as external tables linked to fields of reference, and as absolute and validated number fields (corresponding to an *initial date* and *final date*), in order to define a highly accurate chronological context.

The heterogeneity of the terminology applied to the type of finds was also managed by defining *thesauri*; in this case, they were implemented and managed as external tables linked to the reference fields.

Another problem that arose was the need to place

7 The data collected by various researchers can be compared only by taking into account their intellectual history and individual background (Terrenato 2006:19); data grow old and it would be better to make them available at once, without seeking perfection, when the scientific community is in greater methodological harmony with whoever has produced the data (Gattiglia 2009: 56).

8 The database represents the development and further study of a previous project aimed at implementing the first step towards the development of a GIS for the city of Pisa, resulting in the degree thesis of Francesca Anichini, entitled "Tutela, Ricerca, Valorizzazione del patrimonio archeologico: progetto per il G.I.S. della città di Pisa" (Archaeological heritage protection, research and enhancement: a project of a G.I.S. of the city of Pisa) (Anichini 2004-2005; Anichini, Paribeni 2005). A database was created during the project which contained diachronic archaeological data only and was already based on the minimum spatial unit of the archaeological intervention. The database was developed according to an open structure, making it a "container" of smaller units. Archives suitable for containing the description of the urban tissue were not created; however, given the large amount and heterogeneity of the data, they were collected in a single archiving system for the first time. The questions that guided the implementation of the database were essentially: Where? When? How? What? Easy questions, but necessary for building the foundations of a search that opens the door to many different opportunities and various levels of investigation (Anichini 2004-2005: 85). Practically: the location of finds in Pisa; intervention dating, i.e. the date of execution; quality, that is, the type of intervention; and, finally, the type of find, focusing on two sets of information: chronology and detailed type of find. Chronology was initially divided into wide-ranging historical periods as acknowledged by archaeology which allowed initial overview of the site, diachrony and related potential. More detailed examination was considered for searches focusing on single specialised disciplines. The type of find was covered in the same manner: initial cataloguing with broad classes of belonging and subsequent in-depth examination within the various categories. The system offers different levels of reading and detail; it can be used, therefore, according to different levels of need and interest, and can be quickly and easily consulted when specialised issues need to be considered (Anichini 2004-2005: 87).

9 The decision to choose the Intervention as the minimum unit will also be used for the open data archive. This choice allows data to be updated more easily after the initial entry phase. This is the same criterion used by the Archaeological Data Service of the University of York; for a different position, based on historical topography, see Fronza, Nardini 2009:68

10 The term thesaurus means a list of values that regulate the terms used.

data from documented stratigraphic excavations on the same level as data from occasional information and finds, in order to allow subsequent analysis. While *archaeological intervention* represents the minimum unit of reference (which at the same time corresponds to an identifiable geographical position[11]), Context – with relevant records and material quantification – is the item of evidence with greater detail that needs to be managed in the archiving system. The aim is to make the highest and lowest level of information communicate in the same environment and, therefore, interact in a dynamic process of comparative analysis.

1.1 The logical structure

The problems described above were solved by creating an operating scheme based on four different logical levels that gradually manage information through an interpretative synthesis process. The process, starting from the material traces, transfers the data into typological and chronologically-divided macro categories.

The diagram that synthetically describes the levels and relations between the elements which contribute to defining the information potential of each intervention, starts from Level I which includes the primary data. This level is available only for certain types of intervention and, among these, only those performed recently and for which documentation has been preserved. Primary data are faithfully reported in detailed records in order to reproduce the entire stratigraphic sequence (cfr.§ 5).

Level II of the diagram contains data referring to the interpretative synthesis phases carried out by the authors of the work: description of Groups, Phases and Periods for the excavation data; reports or simple accounts of interventions for which we have no other kind of documentation.

Level III contains Level I and Level II disassembled data. The following are taken into account: general details of intervention, information source, geographical reference, and reliability of data and additional information which allow their description and classification by single find.

It is at this point that we encounter data produced and recorded with very different methods. The researcher must review and interpret the data and actively work on synthesising and classifying the data. The archaeological finds need to be classified in order to standardise them and allow their comparative analysis. While during the first two stages the researcher's work is to catalogue, computerise and partially review the documents, during the third stage the researcher directly converts them into standardised categories of archaeographic and archaeological data. This is a difficult step mainly because of the need to interpret poorly-detailed information which either describes very general material traces or provides an interpreted term without specifying its origin in the material traces. A critical review of the data unavoidably follows, which includes overall analysis of the intervention in terms of period of execution, type, executor's features, etc., which determine an overall level of reliability of the information acquired.

The data standardisation process is structured along four levels of synthesis which allow the information to be analysed according to various levels of investigation in both spatial and conceptual terms (see § 6).

Every trace is gradually related to four categories: the first defines the typological-qualitative component, the second defines the typological-functional component, while the third and fourth categories define the role of the specific datum within a broader system of spatial relations according to two different interpretations: local and urban.

Level IV focuses on analysis: the classified data are compared and can be subject to processes for developing and creating new synthesis information.

2. Software

The software chosen was the proprietary product: Microsoft Access. In order to understand the reasons for this choice, it should be remembered that this research project is the continuation and further evolution of a previous project (Anichini 2004-2005; Anichini, Paribeni 2005), already developed in Windows environment with the same proprietary software. The reasons that brought to this decision are briefly summarised below[12]:

- compatibility with other software, especially GIS software (ESRI ArcGIS 10);
- compatibility with software used by other institutions: Dipartimento di Scienze Archeologiche (Pisa University), Soprintendenza per i Beni Archeologici della Toscana and Comune di Pisa;
- greater knowledge of this software by the research team[13].

3. The RDBMS structure of archaeological data

The RDBMS architecture is based on a series of tables linked to each other. The tables contain the archived data and the *thesauri* or lists of values necessary for

11 Not necessarily univocal; several interventions that have taken place over the years may refer to one location.

12 The discriminating factor when evaluating the best software for creating a DBMS is the software's capability to manage archive complexity and its compatibility with the applications it must communicate with (Gabucci 2005: 32)

13 We decided that it was preferable to develop the project with the software that the majority of research team members could easily use, with a view to optimising DBMS development and implementation with respect to the main research objective and the strict project schedule. Our main aim then (as also now) was not to develop an open-source software specifically dedicated to this type of archaeological archiving.

filling in certain fields in guided mode. The data tables have a user-friendly form, which allows easy entry of data and rapid navigation among the forms. The forms are grouped into two different user interfaces: the first, where the forms show all the fields of the data archiving tables, specifically addresses data entry operators (administrators); the second, where the forms mainly provide query results, specifically addresses users consulting the database.

The tables, called *thesauri*, have accessory features and are used by the RDBMS to standardise the language used. It is essential to normalise the language of a database, especially as regards the synthesis fields, in order to use the data. Non-standardised language can lead to inefficiencies that can also make the collected data completely unusable. The legibility and interpretation of a database highly depends on the formal clarity and completeness of these instruments. Basically, when designing a relational architecture, great effort must be taken to create an efficient database (Fronza 2004: 415). *Thesauri* are divided into:

- closed *thesauri*, with values that cannot be modified by the searcher, referring to dictionaries with a very high level of language processing, such as that used to regulate chronological periods. In some cases, they also include the value other to help the operator in the event of missing items.
- open *thesauri*, which do not apply restrictive control to language and allow the operator to overcome any restrictions by automatically updating when entering data. These thesauri have been adopted for fields for which it is not currently possible to establish a univocal set of values. These dictionaries tend to gradually change into closed lists as the reliability of the sample increases (Fronza 2004: 416). Open *thesauri* can also be used for data that, given their nature, cannot easily be used for closed lists: for example, fields containing the name of the Principal Investigator or the individuals performing the investigation.

The structure is able to simultaneously archive and analyse the data that will be needed for calculating archaeological potential, except for geological and geomorphological data. A specific database and geodatabase have been created, respectively, for these data. This report will exclusively take into account the data archiving structure for buried archaeological heritage, while the solutions adopted for urban data and the building archaeology results will be published over the following months, together with the aerial archaeology data, historical data and mapping data.

Another feature of the database is that it provides access to the digital archive directories and has a simple list-directory function that indicates the physical location of finds and of all external "paper" documentation archives. This system allows ongoing updating of data and offers users the opportunity to know if and/or where to find existing data, and to enter more specific data at a later stage.

3.1. Intervention Record

A record was prepared drawn on the idea that the archaeological intervention is the minimum common denominator, i.e. the minimum item of reference for the topographical management of buried archaeological data. The record highlights the basic features and information by identifying the main characteristics, type and chronological setting of the finds, as well as the source of information. Although referring to the same spatial unit, the guiding principle of the intervention record was to provide the system user with information items that did not have a high degree of synthesis, allowing specific data and further details to be obtained and ensuring reference to specific documentation available (from lesser to greater detail).

The record is divided into sections:

- Technical and topographic data
- Chronological data
- Documentation related to the intervention
- Source of information
- Drafting data

3.1.1 "Topographic and technical" data section

This section includes the technical data regarding the intervention. It provides details on the topographic parameters, methods (both typological and those related to the executors and Principal investigators) and chronology.

Fields:

Intervention ID: "number" field, univocal ID number code for the intervention.

Location: "text" field; this field is more general and less analytical than a common topographic name. We preferred leaving all the indications (either specific or indicative) found in the sources; this method may appear to be inconsistent, however, given the essential nature of the field for mapping purposes, it ensures greater accuracy where possible. Further details, where existing, are entered in: "*Notes*".

Road/square: "text" field linked to the "roads" *thesaurus* containing the official road names of the municipality of Pisa, provided by the municipality.

Site Code: "text" field; this field contains the Site Code assigned to the intervention, if existing. It is used as reference when searching for specific archiving documentation and locating mobile artefacts. This field is also used to generate the univocal code for Context records.

Type of intervention: "text" field linked to the "type of intervention" thesaurus which lists 10 items that define the type of work:

1. *watching brief*
2. *coring*
3. *geophysical/geochemical survey*
4. *occasional finding*
5. *field survey*

6. *rescue excavation*
7. *preventive excavation*
8. *research excavation*
9. *inspection*
10. *not specified*[14]

Intervention Date: "text" field containing, where available, the date of the intervention[15].

Year: "number" field containing only the year of the intervention. For long-term interventions, the start year of the works is reported. This field allows targeted searches as well as data screening thanks to an information reliability parameter related to the period of execution.

Duration of Intervention: "text" field providing preliminary information about the duration of the intervention, expressed in working days (1 month = 20 working days).

Executor of Intervention: "text" field containing the name of the person, team, company or institution who/which materially performed the intervention; this field is linked to an open *thesaurus*[16].

Principal investigator: "text" field containing the name of the Principal investigator(s); this field is linked to an open *thesaurus*.

Number of samples: "text" field; since each excavation area/sample is considered as the minimum reference unit, the overall number of sampled areas within one work[17] is calculated in a specific field. As a result, the number of samples belonging to an investigation and the number of intervention records that have been filled in can be immediately assessed. The number is reported in digits separated by "/"; the first digit is a progressive number that indicates the sample whose respective record is being viewed, while the second digit indicates the overall number of samples performed during the intervention.

Size: "number" field indicating, where available, the square metres of the sample area.

Maximum depth: "number" field reporting the maximum depth (in metres) reached during the excavation.

Depth of Groundwater: "number" field reporting the depth (in metres) at which groundwater was met[18].

3.1.2 "Chronological" data section

This section provides information about the presence of finds (of any type) referring to wide-ranging chronological periods. It is composed of a series of "text" fields, linked to drop-down lists with *yes* and *no* options pertaining to the following periods:

1. *Pre-history*
2. *Protohistoric age*
3. *Etruscan period*
4. *Roman period*
5. *Late Antiquity19*
6. *Early Middle ages*
7. *Late Middle ages*
8. *Modern age*
9. *Contemporary age*
10. *Non identified*

3.1.3 Recording methods section

This section certifies the existence of documentary material related to the intervention. In the event of stratigraphic interventions, indications are only given on the presence or absence of different types of documentation which are separately detailed in the Location Record (cfr.§ 4.3).

A "yes/no" field that must be ticked is available for each type of ***Recording method.***

- *Written records*: includes reports, Context sheets, taphonomic sheets, anthropological sheets, quantification of mobile artefacts and inventory sheets.
- *Photographic records*: includes all general and detailed photographic material.
- *Drawn records*: this field should be ticked only if excavation plans have been preserved (level of detail: context / phase (multi-context) plans or composite plans).

Matrix: "hypertext link" to the file containing the stratigraphic diagram of the intervention.

Folder: "hypertext link" to the folder containing all

14 With respect to the first drafting of the thesaurus, "building archaeology" and "remote sensing" have been removed. These two investigations will be conducted systematically over all the examined area and will be specifically recorded within the RDBMS.

15 In this case, we needed to address different levels of specification, ranging from general periods to specific dates (day, month and year). Since no further information was available and since we believed that it was not strictly necessary for this type of data to be highly specific, we decided fill in the field with information from the source. In cases where the document did not contain this information, if archived by the Superintendency, the year was deduced from the archiving date; instead, in the event of a bibliographical source, we decided not to indicate a date, thus leaving the user to indicate the terminus *ante quem* of the date of publication; in both cases, documents with this problem are considered unreliable for "intervention date" targeted searches.

16 Open *thesaurus* fields were created with a simple expression SQL SELECT DISTINCT [Table]. Field FROM [Table] ORDER BY [Table] [Field].

17 Considered as such when execution times are close and/or consequential and the executor is the same.

18 This value is very approximate. Documents reveal that the point of groundwater surfacing is not systematically calculated. Since probably not considered an important issue, it is available only when water surfacing compromises activities, making stratigraphic readings difficult or forcing excavations to be suspended. There are only few documents that report this value and, of these, many are approximate (Anichini 2004-2005)

19 Intended as the "late Roman" period (see § 6.1)

the documentary material regarding the intervention, preserved in the archive.

Materials no longer traceable: "text" field with drop-down list (yes and no). It reports the presence or absence of mobile artefacts preserved by the Soprintendenza per i Beni Archeologici or other Institutions.

Notes: "text" field that contains details about the information archiving and transmission phase, which are not searchable in the GIS. In addition to the above, this field can also (although rarely) provide clarifications on subsequent fields.

3.1.4 Source of Information Section

This section describes the primary source used for filling in the form, i.e. the source that provides the greatest amount of information. Sources may be archival or bibliographical and, where complementary, both may be present.

Source of Information: "text" field featuring drop-down lists, linked to the "*source of information*" *thesaurus* table which allows the data regarding the body, institution or person holding the information to be entered. *Thesaurus* entries were created for their specific application to the testing area (Pisa):

- *Florence – SBAT*
- *Pisa –SBAPSAE*
- *Pisa – University*
- *Published*
- *Pisa Archivio di Stato – ASP*
- *Pisa Archivio Opera Primaziale*

The following fields were created for the *archival source*:

Protocol number: "text" field used for entering the protocol number provided by the Institution holding the document. This number allows the paper copy of the document and any attachments to be easily traced[20].

Date: "date/time" field that refers to the document protocol date expressed as day, month and year.

Type of document: "text" field that specifies the type of document from which information is taken; it is linked to the "type of document" thesaurus with the following entries:

- *communications*[21]
- *information reports/notices*[22]
- *assignment reports*[23]
- *excavation reports*[24]
- *state of advancement reports* [25]
- *requests for authorisation*[26]
- *collection reports* [27]

Official of reference: "text" field containing the name and surname of the Superintendency official of reference; this field is linked to an open thesaurus.

With regard to data from *published sources*:

Author: "text" field consisting of surname and name initial of the author(s) and/or editor(s) of the publication.

20 For the SBAT Archive of Florence, the two entries are "position" (pos. 9 Pisa 4) and "archiving number". With the recent adoption of the electronic protocol system, only a progressive number is provided in the paper material section instead of geographical indications.

21 Information reports that are related to areas subject to archaeological interest or restrictions: request for information regarding procedures for preventive archaeological interventions, treatment of sites after excavations, communications from institutions regarding the forthcoming presentation of projects (road, sub-services, etc.), material handling, etc.

22 These reports basically relate to occasional recovery; they are drawn up by associations or private citizens who inform the Superintendent or local Official of the location and the circumstances regarding the find/type of material, and request intervention by the Superintendency. Reports containing historical information and technical details about the material are very few.

23 This group mainly includes older documents, drawn up by inspectors or officials from the headquarters of Florence, who were sent on assignments to check reports, verify the progress of excavations, carry out preventive surveys, meet institutions or private citizens to make arrangements for future interventions or for those under way. In some cases, these documents are a sort of brief summary of the state of the art, in other cases they are similar to short excavation reports and provide information about the most important finds (especially mobile artefacts).

24 This Group contains the preliminary reports drawn up by the executors of the intervention or by the Principal Investigator. These reports are an integral part of the excavation documentation delivered to the Superintendency at the end of works. They report intervention timing and methods, the type of find, the size and depth of the area of excavation, and any location of mobile artefacts. In some cases they also enclose plans and graphic reproductions, lists of materials, lists of Contexts and Harris matrix. Photographs are very seldom included.

25 Brief summary by a Superintendency official or by the executor of the intervention on the state of works at the time the document is drawn up. These reports are mostly used for long-term works often in the form of communications to the Superintendent or to the client.

26 Drawn up by an Institution, firm or private citizen, these requests address the Superintendent and describe the project that will be implemented in a protected area or in an area of particular archaeological interest (for Pisa: the area corresponding to the two declaratory judgments of 1986 and 1993). These documents are not of a strictly archaeological nature and are often a useful tool for the exact location of the intervention.

27 This document certifies the collection and/or transfer of mobile artefacts. The site of provenance of the material is indicated as well as the amount of single pieces or number of chests, and the previous location. In the event of delivery by a private citizen, his/her name is mentioned. In some cases, the presence of material is the only documentary evidence of an intervention; it is often possible to trace the area of intervention and year of execution from this type of document and, in more fortunate cases, the general chronological period; of course, any type of spatial, typological and stratigraphic information about the intervention is missing.

Title: "text" field containing the whole title of the article or monographic volume.

In: "text" field, for articles or parts of volumes, reporting the whole title of the volume, review or congress acts in which they are included.

Place/Date: "text" field reporting place and year of publication.

Page: "text" field reporting the pages on which the intervention is mentioned.

3.1.5 Drafting data Section

This section contains data regarding the drafting of records; it allows the name(s) of the compiler(s) to be traced if clarifications are required.

Record compiler: "text" field containing the name and surname of the person who compiled the record; this field is linked to an open thesaurus.

Drafting date: "date/time" field that provides the date on which the record was initially drafted (expressed as day, month and year).

Updating date: "date/time" field provides the date on which the record was most recently drafted (expressed as day, month and year).

Reason for updating: description of the reasons that led to updating the record, compiling further fields or changing the existing ones. Thanks to this field and the previous one, it is possible to update records for work that is still in progress or for work that, once finished, has undergone significant study thus allowing the existing information to be further developed.

4. Records related to the Intervention Record

The table is linked to other 5 tables which specifically take into consideration several sets of information: georeferencing, description, location, bibliography and documentary references.

4.1 Georeferencing Record

A link to a table connects the "topographic and technical data" (§3.1.1) to the georeferencing location of the intervention, through data indicating its reliability and degree of precision, based upon the following fields:

Location ID: "number" field in which the identification number must match the number entered in the InterventionID field of the intervention record.

Acquisition system: "text" field that indicates which system was used for acquiring the coordinates. The following may be chosen from a drop-down list:

- *RD*: Direct acquisition with benchmarks of reference.
- *GPS*: Acquisition through satellite platform.
- *CAR*: Identification using mapping references.

Type of mapping reference: "text" field for cases in which acquisition is via mapping; the type of support used is specified.

- *CTR*: Regional Technical Map.
- *OFC*: regional orthophotomap.
- *CAT*: cadastral maps.
- *IGM*: Military Geographical Institute mapping.
- *IIM*: Marine Hydrographic Institute mapping.

Mapping scale: "text" field containing the mapping scale of reference when the GPS system is not used; a drop-down list provides the following entries:

- *S0K*
- *S1K*
- *S2K*
- *S5K*
- *S10K*
- *S25K*
- *S50K*
- *S100K*

Georeferencing accuracy: "text" field linked to the *"reliability" thesaurus* table; this field provides three levels of scanning accuracy in relation to the georeferencing of each intervention:

- *high*: when obtained through direct acquisition (RD) or using the GPS.
- *good*: when obtained through mapping with scale <= 2K.
- *poor:* when obtained through mapping > 2K.

Site geometry: "text" field that indicates which geometric primitive is used for the site's geometry[28].

Coordinates: the geographic coordinates29 of the polygon centroid representing the area of intervention are reported

- *Z coordinate*: "number" field.
- *X coordinate*: "number" field.
- *Y coordinate*: "number" field.

28 During the data entry phase, a polygonal graphical representation was chosen for GIS vectoring of the single interventions, without determining the type of graph on the basis of the accuracy of perimetration or georeferencing (this decision may be taken during themed visualisation of the datum). The idea at the basis of this decision is to not create point positioning that may lead to errors, creating "false positive" finds in a position that is in any case geographically defined with accuracy because corresponding to a single point. Where information is general and does not allow exact perimetration (e.g.: "find in Via S.Maria") a polygon is drawn which includes the interested area (e.g.: the entire length and width of Via S.Maria) instead of deciding upon a point position arbitrarily or using a centroid.

29 We use Gauss-Boaga coordinates, Monte Mario Italy 1

4.2 Synthetic description record

Given the need to make the Intervention Record more objective, we decided to separate the synthetic description (the result of the researcher's work) from the primary/raw data. The table is composed of the following fields:

Record ID: "number" field in which the ID number must match the number entered in the InterventionID field of the intervention record.

Synthetic description: "Memo" field containing a brief description of the main finds relating to the intervention. The description helps understand and examine the synthetic information entered in the other fields.

4.3 Location Record

This is an essential table for the database because it provides linking to all existing documentation both inside and outside the RDBMS, to digital documentation (inside and/or outside the RDBMS) and to the physical location of the documentation and finds. As in the case of the Recording methods section (infra §3.1.3), it includes written, drawn and photographic records and information regarding mobile artefacts. The table has the following fields:

Record ID: "number" field in which the ID number must match the number entered in the InterventionID field of the intervention record.

Fields related to mobile artefacts:

Location of materials: this free "text" field gives indications on the building/warehouse where the finds are preserved; this field is linked to an open *thesaurus*.

Location of exhibition: this free "text" field reports the museum/exhibition centre that preserves all or part of the finds; this field is linked to an open *thesaurus*.

Internal quantification archive: "yes/no" field that indicates the presence of quantification records inside the database. The table has buttons linking to the internal archive.

External quantification archive: "hyperlink" field, linking to any quantification records existing in the digital archive.

Fields relating to written records:

Location of written records: free "text" field that reports the physical location of written records; this field is linked to an open *thesaurus*.

Internal records archive: "yes/no" field that indicates the presence of written records (Context, etc.) inside the database. The table has buttons linking to the internal archive.

External quantification archive: "hyperlink" field, linking to any written records existing in the digital archive.

Fields relating to drawn records:

Location of drawn records: free "text" field that reports the physical location of the drawn records; this field is linked to an open *thesaurus*

External drawn archive: "hyperlink" field, linking to any files existing in the digital archive.

Fields relating to photographic documentation:

Location of photographic records: free "text" field that reports the physical location of the photographical records; this field is linked to an open *thesaurus*

External photographical archive: "hyperlink" field, linking to any digital files existing in the digital archive.

4.4 Bibliographical Record

Provides details on the bibliography related to the intervention and allows better understanding, beyond the main source of information used for archiving. The table is composed of the following fields:

Record ID: "number" field in which the ID number must match the number entered in the InterventionID field of the intervention record; in the form it appears as a ComboBox originated from the InterventionID field of the intervention record.

Other IDs: "text" field whose alphanumerical content must match the records in the other RDBMS sections.

Author: "text" field containing the surname and name initial of the author and/or editor of the publication; this field is linked to an open *thesaurus*.

Title: "text" field containing the whole title of the monographic volume, part of the volume or of the article; this field is linked to an open *thesaurus*.

In: "text" field, for articles or parts of volumes, reporting the whole title of the volume, review or congress acts in which they are included; this field is linked to an open *thesaurus*.

Year: "number" field, year of publication; this field is linked to the "Year" *thesaurus* table.

Page: "text" field reporting the pages of the article or volume part.

4.5 Documentary Reference Record

This record provides the full digital version of the document used for data entry, allowing direct comparison between the synthetic data and the original source; the following fields are included:

Doc. ID: "number" field (primary key) in which the ID number must match the number entered in the InterventionID field of the intervention record.

Text: "Memo" field in which the full version of the document used for drafting the intervention record[30] is

30 This field was entered in the first RDBMS implemented for Pisa (Anichini 2004/05) and consisted in manually tran-

reported.

Link: "hyperlink" field to the digital file of the document.

5. Excavation records

The RDBMS also contains the excavation records in digital format. These tables wish to faithfully reproduce the archaeological record which progressively describes the stratigraphic sequence of an excavation.

The archive is divided into:

- Tables related to the chronological and interpretation stages of the excavation (Period, Phase, Group);
- Tables related to the stratigraphic data (Context);
- Tables related to the artefacts.

5.1 Context record

This record contains nearly all the items of the ministerial Context paper record sheet with the exception of artefact-related data which are taken from a query between the "Context Record" table and the "Quantifications" table[31]. "Subtype" and "Synthetic interpretation" have been added; these entries are useful for searching in the archive with a standardised common field both as regards material source definition and interpretation. Both fields are linked to open thesauri, whereas the fields "distinction criteria", "method of formation", "consistency" and "status of preservation" are linked to closed *thesauri*[32].

The "Context Record" table is the item required for linking to the Context maps of every excavation in GIS environment and is composed of the following fields:

Context Code: "text" field; this validated field is created from the Site Code and the univocal Context number.

Context: free "text" field (Context number).

Type: ComboBox "text" field with drop-down list (positive or negative).

Intervention ID: free "number" field; this value must be equal to the InterventionID value of the intervention record.

Location: free "text" field.

Site Code: ComboBox "text" field[33].

Code: "text" field, the value entered must match the value entered in the Site Code field

Year: ComboBox "number" field linked to the Thesaurus_Year table.

Area: free "number" field.

Sample: free "number" field.

Sector: free "number" field.

Environment: free "text" field.

Square: free "text" field.

Min. height: free "number".

Max. height: free "number" field.

Maps: free "text" field.

Sections: free "text" field.

Prospects: free "text" field.

Photograph: free "hyperlink" field which links to Context-related photographic documentation stored in the intervention directory.

Archaeological find: "yes/no" field which indicate the presence of Catalog cards.

N: "yes/no" field which relates to the Coin Catalog Card.

Definition and position: free "text" field.

Sub-type: this field connects the different records by synthetically typologising the definition which, instead, must remain a free field in order to conform to the heterogeneous and dissimilar nature of Context and to include all previously-existing definitions that lack standardisation[34].

This ComboBox "text" field is linked to the "sub-type" *thesaurus* and has the following entries:

- *other*
- *opening*
- *hole*
- *erosion/destruction*
- *filling*
- *burial*
- *layer of ashes/bricks*
- *layer of mortar*
- *layer of stones/bricks*
- *layer of soil*
- *wooden structure*
- *horizontal masonry structure*

scribing the documents. Although there was not sufficient time to complete this work in the same manner or with OCR conversion of the scanned documents, it was decided to maintain the record so as not to lose all the previous work and also with a view to future development.

31 The fields corresponding to Masonry Context record entries can be directly filled in. Since a ministerial defined record is only partially available (http://www.iccd.beniculturali.it/index.php?it/251/beni-archeologici) for recording this type of evidence (stratigraphic reading of elevations), we are working alongside experts of this sector who belong to the project team in order to create fields and connected thesauri. The results of this work will be the subject of a future publication.

32 For all Record fields for which specific changes have not been made (except for strictly IT-related issues), please refer to the indications provided by the ICCD on how to fill in Context records http://www.iccd.beniculturali.it/index.php?it/251/beni-archeologici.

33 Generated from the following script SELECT [interventions_Site Code].[Site Code], [interventions_Site Code].[InterventioID] FROM [interventions_Site Code] ORDER BY [Site Code] DESC

34 The goal is to collect all possible types of traces in one of the thesaurus entries. The dictionary was previously tested (Gattiglia 2010) on the medieval finds of Pisa.

- *vertical masonry structure*
- *cut/trench*
- *transformation unit*

Distinction criteria: ComboBox "text" field linked to the "criteria" thesaurus table composed of the following entries: *colour, components, consistency, morphology* and of their possible combinations:

- *Colour*
- *Components*
- *Consistency*
- *Morphology*

Manner of formation: ComboBox "text" field linked to the "formation" *thesaurus* table and composed of the following entries:

- *artificial*
- *progressive-artificial*
- *synchronic-artificial*
- *natural*
- *progressive-natural*
- *synchronic-natural*

There are two fields for the definition of the components:

Inorganic: free "text" field.

Organic: free "text" field.

Consistency: ComboBox "text" field linked to the "consistency" thesaurus table composed of the following entries:

- *cemented*
- *compact*
- *friable*
- *plastic*
- *loose*

Colour: free "text" field.

Measures: free "text" field.

State of preservation: ComboBox "text" field linked to the "alteration" *thesaurus* table composed of the following entries:

- *biological alteration*
- *mechanical alteration*[35]
- *chemical alteration*

Description: free "memo" field.

Equal to: free "text" field.

Contemporary with: free "text" field.

Resting on it: free "text" field.

Rests on: free "text" field.

Covered by: free "text" field.

Covers: free "text" field.

Cut by: free "text" field.

Cuts: free "text" field.

Filled with: free "text" field.

Fills: free "text" field.

Stratigraphically later: free "text" field.

Stratigraphically earlier: free "text" field.

Notes: free "text" field.

Interpretation: free "memo" field.

Synthesis: just as the sub-type field standardises the definition of the Context-related material trace, this field is used for providing a single term for Context interpretation. The field is a ComboBox "text" field linked to the "*synthetic interpretation" thesaurus* table which was prepared in an attempt to synthesise the countless finds already registered in the interventions. The *thesaurus* is open and implementable given the typological and geographical specificity of certain finds.

The table includes the following entries:

- *abandonment*
- *agriculture/vegetable garden*
- *other*
- *cesspit*
- *base*
- *hole*
- *post hole*
- *hole for tree*
- *hole for production activity*
- *hole for hearth of fire*
- *hole for waste*
- *hole for burial*
- *debris/landfill*
- *ashes*
- *colluvium*
- *compression*
- *destruction*
- *"crusta marmorea"*
- *dark earth*
- *wash-out*
- *decay*
- *erosion*
- *overflow*
- *heart of fire*
- *foundation*
- *furnace*
- *melting pit*
- *attendance/use*
- *plaster*
- *hydraulic plaster*
- *wedging*
- *levelling*

35 Also includes percolation.

- *wall*
- *paleosoil*
- *flooring*
- *building site*
- *floor preparation*
- *interface of destruction*
- *rise*
- *backfill of construction cut*
- *backfill/obliteration*
- *stair*
- *carbon layer*
- *layer of mortar*
- *layer of stones/bricks*
- *horizontal masonry structure*
- *vertical masonry structure*
- *cut*
- *construction cut*
- *spoliation*
- *burned soil*
- *roof*
- *trench*
- *wear*
- *loose stone foundation*
- *vault*

Group: free "text" field.

Group Code: free "text" field that links to the related activity record.

Phase: free "text" field.

Phase Code: free "text" field that links to the related phase record.

Period: free "text" field.

Period Code: free "text" field that links to the related period record.

Sampling: free "text" field.

Flotation: free "text" field.

Sieving: free "text" field.

Stratigraphic reliability: free "text" field.

Principal Investigator: free "text" field.

Manager: free "text" field.

5.2 Group record

This table is directly linked to the Context Record and provides a simple description of the activity together with its synthetic definition and dating. All the individual Context records composing it are related to this record. The table is composed of the following fields:

Group ID: free "text" field in which the value must match the Group field value entered in the Context record.

Intervention ID: free "number" field in which the value must match the InterventionID value entered in the intervention record.

Site Code: ComboBox "text" field.

Group code: free "text" field that links the Context record to the related activity record; the value entered, therefore, must match the value of that field.

Definition: free "text" field.

Description: free "memo" field.

Synthetic description: as in the case of the Context Record, this field provides a single, synthetic and standardised term for activity interpretation, thus making the search for stratigraphic sequences in the archive easier. For instance, it may be necessary to search directly in the Group record without any need for Context in cases where a combined approach, instead of a single material trace[36], is able to provide a more complete definition. The field is a ComboBox "text" field linked to the "synthetic Group" *thesaurus* table and corresponds to level IV of the database's interpretative field (cfr.§6)[37].

Initial date: free "number" field.

Final date: free "number" field.

Phase code: free "text" field that links to the related phase record; the value entered, therefore, must match the value of that field.

Period code: free "text" field that links to the related period record; the value entered, therefore, must match the value of that field.

Min. depth: free "number" field.

Max. depth: free "number" field.

5.3 Phase record

This table is directly connected to the Context Record, as the previous one, and gives a simple description of the phase as well as its synthetic definition and dating. All the groups composing the record are connected to it. A synthesis field was not included because an effort of this kind was considered unproductive at this level of the interpretative process, where the definitions are strictly limited to the circumstance of the find and may have a very wide scope. The terms of comparison are of a chronological nature and can be analysed through numerical searching in the dating fields.

The table is composed of the following fields:

Phase ID: free "text" field; the value must match the value entered in the Phase field of the Context re-

36 An obvious example is when Context is simply interpreted as a "layer of coal" which generates the definition of "furnace" only during the group definition phase, combined with other traces. Searching for the definition of "furnace" in the Context records will not lead to any results, instead, a cross search with the Group records will confirm the presence of a furnace, check the relevant interpretation and so trace all Contexts interpreted in this way.

37 Since being an open and gradually implementable dictionary, this is an ongoing phase with data-entry operating tests that are still under way. The complete thesaurus will be published at the end of the work.

cord.

Intervention ID: free "number" field; the value must match the InterventionID value entered in the intervention record.

Site Code: ComboBox "text" field.

Phase code: free "text" field that links the Context record to the related phase record; the value entered, therefore, must match the value of that field.

Definition: free "text" field.

Description: free "memo" field.

Initial date: free "number" field.

Final date: free "number" field.

Period: free "text" field in which the value must match the Period value entered in the Context record.

Period ID: free "text" field that links to the related period record; the value entered, therefore, must match the value of that field.

Min. depth: free "number" field.

Max. depth: free "number" field.

5.4 Period record

This table is directly connected to the Context Record, as the previous one, and gives a simple description of the period as well as its synthetic definition and dating. All the phases composing the record are connected to it.

Period ID: free "text" field in which the value must match the Period value entered in the Context record.

Intervention ID: free "number" field in which the value must match Intervention ID value entered in the intervention record.

Site Code: ComboBox "text" field.

Phase code: free "text" field that links the Context record to the related Period record; the value entered, therefore, must match the value in that field.

Definition: free "text" field.

Description: free "memo" field.

Initial date: free "number" field.

Final date: free "number" field.

Min. depth: free "number" field.

Max. depth: free "number" field.

5.5 Quantification record

This table quantifies the ceramic materials found in the single Contexts. Type is the main field, in addition to function, production, shape, decoration, initial date and final date (chronological range that can be used for numerical searching).

Context code: free "text" field in which the value must match the Context_Code value entered in the Context record.

Context ID: free "number" field.

Intervention ID: free "number" field in which the value must match the InterventionID value entered in the intervention record.

Site Code: ComboBox "text" field.

Type: free "text" field that is linked to an open thesaurus. This choice allows the database to be used in other geographical areas other than the project test area. It is well known that ceramics, especially in certain periods, may belong to classes of specific regional, micro-regional and local areas.

Function: free "text" field that is linked to an open *thesaurus* and gradually implemented. It currently includes the following items:

- *Covering*
- *Cooking utensils*
- *Cooking utensils/storage*
- *Storage/transport*
- *Building*
- *Smoke*
- *Casting*
- *Hydraulics*
- *Insulation*
- *Table utensils*
- *Table utensils/storage*
- *Heating*
- *Textile*
- *Transport*
- *Transport/storage*
- *Transport/miscellaneous*
- *Miscellaneous*

Production: free "text" field that is linked to an open thesaurus. The considerations made for the "Type" field may also be made here.

Shape: free "text" field linked to an open thesaurus and gradually implementable. Data entry is currently being performed.

Decorations: free "text" field that is linked to an open *thesaurus*. It presents the same problems of the "class" field to which it is strictly related.

Number of sherds: free "number" field that indicates the number of sherds for each ceramic class.

Initial date: free "number" field.

Final date: free "number" field.

The two "date" fields provide a range of specific reference for each ceramic class. Chronological searches can be carried out either by comparing different classes or within one class.

The quantification record is linked to the "C***ontext_Dating***" record that allows automatic dating of the Context based upon the dates of the material classes available. This dating, however, is not binding and is proposed to the operator who can decide whether to accept or change the proposed dating.

Context_Code: free "text" field in which the value must match the Context_Code value entered in the

Quantification record.

Context No.: free "number" field.

Initial date proposed: ComboBox "number" field generated from the following script: =DMax("[initial date]";"quantification_class";"[CodUS]=[quantification_class]![Context_Code]"). The automatic calculation considers the greater date[38] among all initial dates[39] (initial date field of the quantification record) of the ceramic classes included in a specific Context.

Final date proposed: ComboBox "number" field generated from the following script: =DMax("[initial date]";"quantification_class";"[Context_Code]=[quantification_class]![Context_Code]"). The automatic calculation considers the greater date among all final dates[40] (final date field of the quantification record) of the ceramic classes included in a specific Context.

Initial date operator: free "number" field. The initial date automatically proposed is confirmed or modified.

Final date operator: free "number" field. The final date automatically proposed is confirmed or modified.

6. Level of interpretative synthesis

Gradual in-depth examination of the information acquired from the system is developed in the "interpretative synthesis record" and in the "level IV record". The former is related to the Intervention record, while the latter is directly connected to the former. Both records create a common environment through which it is possible to compare data from different types of documentation according to the logical layout initially described (cfr.§ 1.1).

Every find recorded in an intervention is described over four standardised definition levels which specify both chronological information and typological-qualitative information.

6.1 Chronology

Chronological data management is by no means a side issue in an archaeological RDBMS. The data provided, in fact, are not only heterogeneous in terms of acquisition method and quality, but also dissimilar in terms of chronological definition. Indeed, it is not always possible to check the dates attributed in the light of new knowledge; they must be accepted, therefore, with a certain margin of doubt. Sometimes data collected recently also present a certain margin of uncertainty due to specific contexts or, more in general, the status of the research. We decided to use a simple system for managing this information, based on a chronological interval defined by two different number fields called "final chronology" and "initial chronology", in which the absolute date is included[41]. We decided to date centuries starting from year 1 and ending in subsequent year 10042. Data is recovered thanks to a query conducted on both numerical data in order to define an interval.

Number scanning is accompanied by a more general chronology expressed in text form which allows simplified searching: the macro-periods, already described in the intervention record, are divided into sub-periods, through a ComboBox "text" field called "**Chronology**" that is linked to the *Chronology thesaurus*[43] table and composed of the following entries:

- *Prehistory,*
- *Palaeolithic,*
- *Upper Palaeolithic,*
- *Middle Palaeolithic,*
- *Lower Palaeolithic,*
- *Mesolithic,*
- *Neolithic,*
- *Early Neolithic,*
- *Evolved Neolithic,*
- *Late Neolithic,*
- *Eneolithic,*
- *Early Eneolithic,*
- *Late Eneolithic,*
- *Bronze Age,*
- *Early Bronze Age,*
- *Middle Bronze Age,*
- *Late Bronze Age,*
- *Final Bronze Age,*
- *Iron Age,*
- *Iron Age I,*
- *Iron Age II,*
- *Etruscan Period,*
- *Etruscan Orientalizing Period,*
- *Etruscan Archaic Period,*
- *Etruscan Classical Period,*
- *Etruscan Hellenistic Period,*
- *Roman Period,*
- *Mid-Republican Roman Period,*
- *Late-RepublicanRoman Period,*

38 Since this is a number field, greater date means the most recent date.

39 Initial date means the oldest date of the range.

40 Final date means the most recent date of the range.

41 According to some, the heterogeneity and uncertainty of chronological data is particularly evident in the accurate and precise archaeological information entered in the GIS; furthermore, the uncertainty resulting from an inaccurate chronology leads to the need to incorporate this level of indecision within the data architecture itself (Harris, Lock 1995).

42 For chronologies "before Christ", the numbers will be preceded by a minus (-) sign.

43 This thesaurus was taken (and modified) from the Guidelines for Drawing up the Archaeological Map of Tuscany (Francovich, Pellicanò, Pasquinucci 2001: 195) and from (Anichini 2004-2005).

- *Imperial Roman Period,*
- *Early Imperial Roman Period,*
- *Mid Imperial Roman Period,*
- *Late Roman Period,*
- *Early Middle Ages,*
- *Early Middle Ages VII-VIII century,*
- *Early Middle Ages IX-X century,*
- *Late Middle Ages,*
- *Late Middle Ages XI-XIII century,*
- *Late Middle Ages XIV-XV century,*
- *Modern Age,*
- *Modern Age XVI century,*
- *Modern Age XVII century,*
- *Modern Age XVIII century,*
- *Contemporary Age,*
- *Contemporary Age XIX century,*
- *Contemporary Age XX century,*
- *Not determinable.*

In cases where information does not allow more specific details, the *thesaurus* proposes a standard division where it is always possible to enter a more general item (which matches the macro-period definition, e.g.: Roman, Etruscan, Modern Age, etc.). Where possible, the sub-periods have chronological ranges expressed in centuries (e.g.: Modern Age XVII century); regarding the other periods, we listed the names used in different cultural environments (e.g.: Second Iron Age or Etruscan Hellenistic Period)[44], although specific names are used for our area of study (e.g.: Hellenistic Etruscan Age will be used instead of Second Iron Age). The number field avoids this problem and makes all data immediately comparable.

6.2. Synthesising finds: from material traces to the city

By developing the steps that steered the archive's logical structure (§1.1), each find referable to a specific chronology is progressively related to four definitions of synthesis.

6.2.1 Level IV Record

The first level of synthesis, called Level IV, establishes the type and quality of the find. The following fields are included in the record:

Intervention ID: free "number" field in which the value must match the InterventionID value entered in the intervention record.

Period ID: free "text" field that corresponds to the period code and to the ID number of the intervention (e.g.: intervention no. 1, Late Medieval Age = LM1)

Level IV: free "text" field that links to an open thesaurus and is gradually implementable. The definition adopted defines the typological-qualitative features of the find. Although relating several categories to a single term that contains the main specifications[45] may seem an easy task, the large variety of different and more or less complex material evidence implies that the pre-established dictionary must be implemented with the most common terms[46].

Chronology: "text" field linked to the "chronology *thesaurus*" field (§ 6.1) that lists the sub-periods.

Initial date: free "number" field. This field (as in the case of other date fields) shows the earliest date attributable to the origin of the material trace.

Final date: free "number" field that indicates the most recent date attributable to the material trace.

A *"reliability"* field is also included regarding the reliability of the trace and its interpretation. The criteria used for assigning the reliability value will be discussed in the next working phase, when all the data entered will be reviewed and checked by archaeologists specialised in the different historical periods.

6.2.2 Level III Synthesis Record

This Synthesis record, which may be accessed directly from the Intervention Record form, contains **Level III, II and I**[47] definitions. While the user's interpreta-

44 The Hellenistic Period is the period from the death of Alexander the Great (323 BC) to the battle of Actium (31 BC) where Octavian defeated Anthony and Cleopatra and gave way to a new political course. Pisa had already entered the Roman influence in III century BC, yet it received Roman citizenship only after the social war (89 BC) and after becoming a Roman municipality. From this date onwards, therefore, the correct term to be used for Pisa is the Late Republican Period, not the Hellenistic Period (nor the "Mid-Republican Roman Age"). We decided to extend this period up to 28 BC, since the following year Octavian received the title of Augustus and established the Principate. The Early Imperial Period coincides with the Julio-Claudian dynasty, up to 68 AD (death of Nero), while the Mid Imperial Period is the period from 69 to 192 AD (death of Commodus). Under this emperor's dynasty, the complex economic transformations that had steered the empire's production from the Italic peninsula to Gaul and Spain, and then to Africa, came to an end. The Late Roman Period or Late Antiquity started in 193 AD and ended with the arrival of the Lombards. The exact date in which Pisa was conquered by the Lombards is uncertain, so it was decided to arbitrarily set the end of this period as 600 AD.

45 The term "ceramics", for example, refers to both the type (mobile artefact) and quality (ceramics).

46 It must be remembered that certain traces are specific only to certain areas and/or cultures. During this phase, each operator can implement the thesaurus, after trying to match the find to the list; this avoids the creation of an enormous amount of entries with very slight and unimportant differences. This activity leads to an intense methodological debate in the working team, aimed at reflecting on the actual archaeological meaning of each term: the absence of standardised definitions reveals how archaeologists coming from different backgrounds (especially chronological specialisations) use the same term for different traces, or vice versa, use different terms for the same trace. We believe that it will be possible to closely and thoroughly examine this issue only at the end of this phase, when the sample of data will be more consistent.

47 The relevant *thesauri* will be the subject of a specific report.

tive process in Level IV definition simply consisted in finding a single term for all the typological-qualitative features of the find, in the following three levels, the process becomes a substantial part of the description of the data and of their relations with space. The typological-functional (Level III) features are gradually defined as well as the role taken up by a specific record in relation to space, on both local and urban grounds (Level II and I).

Level I, II and III fields are ComboBox "text" fields. Each field links to its own "*thesaurus*" table which, apart from Level I, is influenced by the choice options of the term entered in the field of the previous level[48].

The reference period is entered according to a chronological range (as in the Level IV record): **Initial Date** and **Final Date**. The fields can be expressed either in text form, by choosing a sub-period ["text" field linked to the "chronology thesaurus" table (§ 6.1) that lists the various sub-periods], or in numbers [free "number" field; see described criteria (§6.1)]

The following identification and linking fields are available:

Intervention ID: free "number" field in which the value must match the Intervention ID value entered in the intervention record.

Period ID: free "text" field that corresponds to the period acronym and to the ID number of the intervention (e.g.: intervention no. 1, Late Medieval Age = LM1)

Height: free "number" field. Where existing, reference to height expressed as metres above sea level can be associated to every find.

Reliability of height: the height of certain finds is essential for developing period DTMs. As already mentioned, however, depth-related data reported in documentation are mostly relative, referring to vague ground levels or just as vague relative 0 points, and are only very seldom absolute. Checking existing altimetry data and making them absolute, i.e. referring them to sea level, is essential. The data obtained, however, are influenced by the greater or lesser indefiniteness of the reference chosen and, therefore, are not equally accurate. For this reason, we decided to support depth-related data by checking their reliability along a triple-scale of values: *exact, good calculated reliability and poor calculated reliability*. The procedure for making height data absolute was conducted in GIS environment by comparing the relative heights to the 3D land model and subtracting the relative value indicated in the documentation. The degree of reliability of the calculated height– *good calculated reliability or poor calculated reliability* – was attributed depending on how accurate the relative 0 point was. The exact value was attributed for data that already had an absolute reference point since directly chosen during the excavation phase.

This is a ComboBox "text" field linked to the " *reliability thesaurus*" table composed of the following entries:

Exact: absolute height a.s.l. directly measured by the person conducting the excavation.

Good calculation: absolute height calculated a posteriori for which it is possible to determine the Ø point of origin of the height (Ø point error <= 10 cm) with good accuracy.

Poor calculation: absolute height calculated a posteriori for which it is not possible to determine the Ø point of origin of the height (Ø point error > 10 cm) with good accuracy.

Height missing: it is not possible to calculate the absolute height because the relative height or the relative Ø point are missing, or because it is not possible to position the Ø point of origin of the height.

6.2.2.1 Evaluating the reliability of record categorisation[49]

The synthesis process developed in the four levels described, transforms primary archaeological information into standardised categories which directly contribute to the calculation of the archaeological potential. This is a very delicate journey because it has a direct impact on the analyses and on the historical-archaeological and mathematical considerations that lead to the achievement of the final project product. We have often underlined in this report how difficulties such as the inconsistency, shortage and absence of data recorded in the sources have unavoidably influenced this process. Although we have designed an archiving system that attempts to reduce these problems as much as possible, we believe it essential to include a field that evaluates, with clear parameters, the overall reliability of data categorisation.

Since the source, intended as the archaeographic documentation of the intervention, was considered the crucial issue, we created a table for evaluating the documentation of every intervention. The starting point was based on the assumption that the greater the amount of documentation, the greater the reliability of the information. Complete documentation allows more accurate checking and understanding of the raw data, as well as greater reliability of its categorisation. Although our aim was not to achieve a qualitative evaluation of the documentation (in the absence of acknowledged standards that exactly define what documentation should contain, but most of all with which criteria should be used for drafting "quality" documentation), a parameter of this kind necessarily had to be used alongside the strictly quantitative parameter.

48 Regarding the "Level II" field, only the entries included in the definition chosen for Level I can be chosen. Regarding the "Level III" field, the available terms are inside closed thesauri according to the entry chosen in Level II. The principle is that every level is more detailed than the previous one, if taken from Level I to III, or more synthetic, if taken from Level III to Level I.

49 We would like to thank the following project team members for their contributions to this problem and for drafting the first report, upon which this paragraph is based: Antonio Campus, Lorenza La Rosa, Claudia Sciuto and Giulio Tarantino.

The result of this work is a calculation table that assigns a score to each intervention, according to a set of simple and previously-defined simple steps, allowing the judgment parameters used by operators to be as coded and objective as possible. The main fields are the type of intervention and type of documentation.

In order to simplify any existing heterogeneity, the different types of intervention were grouped into three macro-categories sharing similar features in terms of type of documentation and information potential:

1. *Geognostic surveys*: coring
2. *Surface surveys*: surface investigations, geo-physical surveys
3. *Excavations*: stratigraphic excavations (all types), assistance, digging, trenching, etc.

At the same time, different types of documentation were identified that need to be produced for every intervention according to standards (more or less recognised standards, yet in any case indicated by ICCD – Central Institute for Cataloguing and Documentation); in this case also, information must be grouped into categories that allow documents acquired or drawn up with different methods to be entered:

1. *Written records*
2. *Photographic records*
3. *Drawn records*
4. *Reporting*
5. *Quantification of finds*
6. *Matrix*

Every category of documentation is interpolated with the three macro categories, then three levels of accuracy are defined for each interpolation regarding the completeness and accuracy in drafting the documentation. The three levels are the following:

- Level one: absence of documentation, value = 0
- Level two: intermediate, value = 1
- Level three: highest information level, value = 2

The sum of all the scores acquired for each category of documentation generates an overall value of reliability of the information included in the archaeological record.

The parameters used for evaluating the type of intervention in every category of documentation are described below.

Given the nature of the type of intervention they refer to, some interventions cannot attain the highest information level (level III) in certain categories of documentation. In these cases, it was established that the intermediate level corresponds to the greatest information potential that may be achieved and documented from that intervention.

Level III automatically includes Level II entries.

Written records

Geognostic surveys:

Level I: the core does not contain any type of Context recording.

Level II: description of Contexts identified in the core with description of most important features.

Surface surveys:

Level I: no evidence recorded.

Level II: Topographic Unit (TU) sheets[50] compiled.

Excavations:

Level I: no Context records.

Level II: the complete list of Contexts is available, some Context records have been filled in, or the majority of Context records have been filled in but not completely.

Level III: all or almost all the Context records have been filled in, at least the main fields (general data and reports) according to ICCD (Central Institute for Cataloguing and Documentation) indications.

Photographic records

Geognostic surveys:

Level I: the core does not contain any photographs.

Level II: general photos (intervention area, boxes) and photographs of the stratigraphic sequence identified in the core.

Surface surveys:

Level I: there are no photographs.

Level II: photographs of identified lands or of specific evidence, with name of TUs and appropriate references.

Excavations:

Level I: there are no photographs.

Level II: general photographs, sporadic photographs without Context documentation, illegible photographs (out of focus or bad lighting) or photographs without metrical or orientation references.

Level III: photographs of all Contexts and any specific items, or highly legible photographs (high definition) with correct metrical and orientation references.

Drawn records

Geognostic surveys:

Level I: the core does not contain any type of drawn records.

Level II: drawn rendering of the stratigraphic section (LOG) with indication and characterisation of identified Contexts.

Surface surveys:

Level I: no drawn records.

Level II: mapping and delimitation of identified TUs.

Excavations:

Level I: no drawn records have been produced or only free sketches are available; the plans do not have any type of reference (benchmark) that can link them to the area of excavation or among one another.

50 Comparable to Site Sheets

Level II: only composite plans or phase plans are available, where the limits of the single Contexts are not identifiable; heights are not reported or are reported inadequately.

Level III: all Context plans are available with correct heights (absolute heights or relative heights with 0 point known) and benchmarks of reference.

Reporting

Geognostic surveys:

Level I: there are no reports.

Level II: a general report is available that simply summarises the interpretation of the evidence, without any reference or description of raw data.

Level III: complete report with description of the stratigraphic sequence and single Contexts.

Surface surveys:

Level I: there are no reports.

Level II: a general report is available that simply summarises the interpretation of the evidence, without any reference or description of raw data.

Level III: complete report with description of primary data.

Excavations:

Level I: there are no reports.

Level II: non periodised report, without description of the material source or with illegible excavation diaries or reports (syntax errors).

Level III: complete ohasing report accurately referring to the raw data.

Quantification of finds

Geognostic surveys:

Level I: the material has not been studied.

Level II: any material found in the core has been catalogued.

Surface surveys:

Level I: the material has not been studied.

Level II: only little information is available about the material found, with partial documentation (only a few photographs, drawings or quantification tables).

Level III: washed/initialled material, photographs of all material per Context, drawings and quantification tables.

Excavations:

Level I: no drawings, photographs and quantification tables.

Level II: only little information is available about the material found, with partial documentation (only a few photographs, drawings or quantification tables).

Level III: washed/initialled material, photographs of all material per Context, drawings and quantification tables.

Matrix

Excavations:

Level I: not performed or illegible.

Level II: existing and performed according to agreements shared by the scientific community[51].

7. User interface

A specific part of the database (starting from the opening screen) has been created for users who are not administrators. Users can consult and search the database through a number of forms, but cannot change or enter data. The user interface was conceived to make the archaeological data easier to read; for this reason, each form contains several tables. The user interface forms were created by means of queries between different tables and of sub-forms. The "intervention record" user form, besides providing intervention table fields, also contains the georeferencing coordinates (from the localisation table), the synthetic description (from the synthetic description table) and the records related to the intervention according to the four levels of categorisation. The user immediately has an overall view of every single intervention and can carry out many search activities (chronological and typological) since all visible fields can be examined. The "sequence" button connects to the "sequence" query form which describes the overall sequence of an excavation thanks to intervention-related data from the categorisation level tables, the period table and the phase table. From these tables it is possible to reach the single phase or period records which include respective phases and activities, and finally the Context record, which is presented as a shorter version compared to the complete record and also contains the quantification data of the finds.

51 It is necessary to consider this specification because matrices have been found that, although performed very recently, did not comply with the Harris diagram rules, but with non-specified rules that do not refer to any known bibliographic reference; furthermore, they do not have an sort of key to help understand symbols that are, at the moment, incomprehensible.

Bibliography

Anichini F., 2004-2005, *Tutela, ricerca, valorizzazione del patrimonio archeologico: progetto per il G.I.S. della città di Pisa*, t.d.l. Università di Pisa, rel. prof. M. Milanese

Anichini F., Paribeni E. 2005, *Il Gis Archeologico della città di Pisa. Primi risultati per la tutela e la gestione del patrimonio archeologico pisano*, «Notiziario della Soprintendenza Archeologica per la Toscana»,1, pp. 205-210

Bagg J, Ryan N. 1997, *Modelling historical changes in Southern Corsica*, in Kemp Z. (a cura di), *Innovations in GIS, London*, pp. 42-55

http://kar.kent.ac.uk/21508/1/Modelling_historical_change_in_southern_Corsica.htm

Bini D., Dubbini N., Steffè S. 2011, *Modelli matematici per la determinazione del potenziale archeologico*, in MapPapers 4-I, pp.68-76

http://mappaproject.arch.unipi.it/wp-content/uploads/2011/08/MapPapers_41.pdf

D'Andrea A. 2006, *Documentazione archeologica, Standard e trattamento Informatico*, Budapest

http://docenti2.unior.it/doc_db/doc_obj_18122_25-03-2011_4d8c69c14d60e.pdf

Francovich R., Pellicanò A., Pasquinucci M. (a cura di) 2001, *La Carta Archeologica. Fra ricerca e pianificazione territoriale. Atti del Seminario di studi organizzato dalla Regione Toscana Dipartimento delle Politiche Formative e dei Beni Culturali*, Firenze

Fronza V. 2003, *Principi di database management in archeologia: l'esperienza senese*, in Fiorillo R., Peduto P. (a cura di), III Congresso Nazionale di Archeologia Medievale (Complesso di Santa Sofia, Salerno, 2 - 5 ottobre 2003), Firenze, pp. 629-632

http://unisi.academia.edu/vittoriofronza/papers/92225/2003-principi_di_database_management_in_archeologia_lesperienza_senese

Fronza V. 2004, *Database Management applicato all'archeologia nell'ambito del progetto "Paesaggi Medievali"* in Relazione del progetto archeologia dei paesaggi medievali (anno 2000 – 2004) pp.399-451

http://www.paesaggimedievali.it/

Fronza V. 2009a, *L'archiviazione del dato in archeologia*, in Fronza V., Nardini A., Valenti M. (a cura di) 2009, *Informatica e Archeologia Medievale. L'esperienza senese*, Firenze, pp. 29-43

Fronza, Nardini 2009, *Un sistema informativo per la gestione dello spazio urbano*, in Volpe G., Favia P., V Congresso Nazionale di Archeologia Medievale, Foggia-Manfredonia 30 settembre-3 ottobre 2009, Firenze

http://unisi.academia.edu/VittorioFronza/Papers/125095/2009_-_Un_sistema_Informativo_per_la_gestione_dello_spazio_urbano

Gabucci A. 2005, *Informatica applicata all'archeologia*, Roma

Gattiglia G. 2009, *Open digital archives in archaeology a good practice*, in Archeologia e Calcolatori, Supplemento 2, 2009, pp. 49-63

Gattiglia G. 2010, *Pisa tra VII e XIV secolo alla luce dell'archeologia*. Tesi di dottorato, Pisa 2010

Gattiglia G. 2011, *Pisa nel Medioevo*, Pisa

Harris T.M., Lock G. 1995, *Toward an evaluation of GIS in European Archaeology: the past, present and future of theory and applications*, in Lock G., Stancic Z. (a cura di) 1995, *Archaeological and Geographical Information Systems: a European Perspective*, London, pp. 349-365