

Spatial Data Management Aspects in Archaeological Excavation Documentation¹

Dieter Pfoser¹, Thanasis Hadzilacos¹, Nikos Faradouris², Kriton Kyrimis²,

¹ Research Academic Computer Technology Institute, Davaki 10, 11526 Ampelokipoi - Athens, Greece

{pfoser, thh}@cti.gr

² Talent SA, Karytsi Sq. 4a, 105 61 Athens, Greece

{niko, kyrimis}@cti.gr

Abstract. Archaeological excavation documentation poses several challenges to spatial data management. This work gives a data model, respective query functionality and the description of the resulting prototype system for recording and analyzing the data produced during the course of an excavation. Based on a requirements analysis, we develop a data model that not only captures geo-referenced finds (pottery sherds, bones, etc.) and their properties, but also excavation diaries and the particularities of the excavation space. Towards an effort of analyzing the collected data and constructing a virtual excavation space, we describe spatial query capabilities providing access to the data in two and three-dimensional space. The system also captures uncertainty with respect to position, which is quite common due to either missing information and/or approximate measurements. Finally, we will give a complete overview of “Arxaiorama,” the system prototype that was developed during the course of a research project and that is currently being used at the excavation site of Dispilio in Kastoria, Greece.

Keywords: We would like to encourage you to list your keywords in this section.

1 Introduction

Excavation documentation is an intriguing subject when confronted with the vastly varying needs of the users, the archaeologists. The basic requirements to a tool supporting such a task are the holistic comprehension, management and promotion of the results of excavational by providing (i) flexible import of data, i.e., location-independent entering of data through Web clients, (ii) dynamic visualization of the excavation space, i.e., use of query functionality to visualize the collected information in 2D and 3D, and (iii) reporting and documentation of the excavational progress by means of printed and electronic reports.

¹ Other contributors: Manolis Koutlis, Talent SA, Greece, Dora Nousia, Research Academic Computer Technology Institute, Greece, Marina Sofronidou and George Chourmouziadis, Aristotle University of Thessaloniki, Greece.

Based on these requirements the prototypical excavation documentation system Archaiorama was developed. It provides a graphical user interface for data collection, data analysis, and reporting and is built on top of a PostgreSQL DBMS [10] implementing a data model specifically designed for this application context. Archaiorama exhibits a strong spatial data management component and, given its query functionality, visualization capabilities and data model, can be considered a custom *geographic information system for excavation documentation*.

This work presents the spatial aspects of the overall system as related to the data model, query capabilities and emerging algorithmic issues related to data management. The Archaiorama system is an ongoing effort documented in previous publications [2] [6] and system prototypes [3].

Currently a large number of applications exist that address partially the requirements as outlined above. Typically these application are derived from CAD software, e.g., ArchaeoDATA [1] and Singularch [11] and provide an extensive functionality with respect to the spatial data aspect of excavation documentation including interfaces to respective instruments. In contrast, the objective for Archaiorama was to provide a custom light-weight system that provides only the necessary functionality and can be used as a server application providing location independent access. While custom solutions for data models exist, e.g., [6], there is not consistent treatment of data modeling for excavation documentation. A related effort towards a coherent data model for cultural heritage information is CIDOC Conceptual Reference Model (“CRM”) [7], a formal ontology intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information. However, CRM represents an abstract model that is not directly useful for the definition of a specific data model for excavation documentation.

The remainder of this paper is organized as follows. Section 2 presents a general system overview of Archaiorama. Section 3 describes the excavational finds, their (spatial) properties, particularities of the excavation space, and the spatial aspects of the data model. Section 4 outlines query capabilities and visualization options and, finally, Section 5 gives conclusions and directions for future research.

2 Archaiorama Overview

The objective for the development of Archaiorama is to provide a tool for the holistic comprehension, management and promotion of the results of excavational work. In terms of a data management this task relates to *data collection*, *data analysis* and *reporting* (cf. Fig. 1). The system requirements were developed in cooperation with the archaeological team conducting the excavations in Dispilio, Greece [5].

Archaeological excavation is a process that produces a large amount of data in the form of *finds*. Finds represent spatial objects, since their position in space matters [4]. Moreover, we are dealing with three-dimensional spatial objects, since besides longitude and latitude, also the depth of a discovery is relevant. All relevant attributes of the objects we want to keep in our system have to be captured.

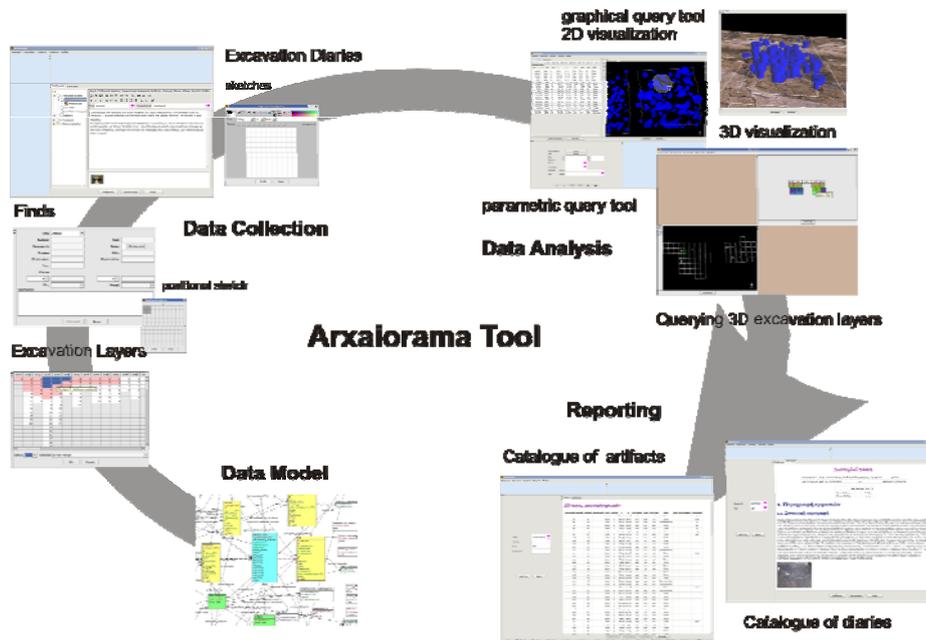


Fig. 1. Excavation documentation system “Arxaiorama”: system overview

Another data component is *excavation diaries*. Diaries capture the progress of an excavation. Importantly, the circumstances as to where and how an artifact was discovered are also kept in the diary. In our case diaries follow a certain layout and structure but are typically composed of free text and contain pictures. The system provides respective forms for both data components.

The data collection for finds and diaries is supported by an editor that requires the user to complete specific forms depending on the type of information that is entered in the system (cf. Fig. 1, upper-left corner, “Data Collection”). Data entry is also supported by integrated tools such as (i) a drawing tool to *sketch* specific excavation constellations and/or large finds and (ii) to record the position of larger finds by means of a positional sketch (cf. Section 3.2). The data is stored by means of a PostgreSQL database [10] using PostGIS extensions [9]. A respective data model was defined in co-operation with the domain experts, the archaeologists.

Once the data is collected in the system, Arxaiorama provides a holistic view of the excavation data, i.e., switching from an atomic view of excavated items to an integrated representation of data based on space that permits further data analysis. With space being the underlying means to structure information, special emphasis is given to spatial query capabilities, which are as follows.

- 2D graphical user interface for spatial range queries (cf. Section 4.1)
- 3D visualization and navigation of query results and object identification (cf. Section 4.2)
- exploration and structuring of excavation space (visualization and querying of layers, grouping into conceptual layers (cf. Section 4.3).

Finally, the reporting feature of Archaiorama allows for exporting (i) any query result (finds) and (ii) any diary in html format. This feature allows for easy viewing of such results with a Web browser, but also the easy integration of such reports in existing Web sites.

3 Data and Data Model

An important aspect in the overall system design is the creation of an adequate data model. While the complete Archaiorama data model comprises 42 entities and 56 relationships, we, in the following, describe only its spatial aspect by surveying the data and the particularities of the excavation space to finally describe the data model itself.

3.1 Excavational Data

The conceptual modeling of the underlying database is based on identifying the entities that constitute excavational data, their attributes and the relationships between these entities. In brief, the essential entities are:

- **Diary** (entries): the basic unit of recording the excavation progress and work. Each day, for each section that was excavated that day, a diary entry is created, in which the work progress and the finds are recorded. Besides text, photographs and sketches are used to document the trench and/or finds.
- **Finds** and their various categories: anything remarkable that was found ranging from complete vases, housing outlines, to grains of pollen. The following major categories can be initially identified given the context of the specific excavation of Dispilio, Greece [5]: (i) *soil samples*, (ii) *postholes* (with wood), (iii) *pottery* (sherds), (iv) *small finds* (bone fragments, tools (iron, stone), and (v) *large finds* (stoves, fireplaces, walls).
- **Trenches**: to ease the management of the overall process, the excavation space is subdivided in a set of regular cells, in the case of this specific excavation 5x5m. A cell is referred to as trench.
- **Layers**: parts of a trench that were excavated during the course of several days and present common characteristics are characterized as a *layer*. This characterization is on a per-trench basis and depending on the experience of the archaeologist in charge can be more or less conservative, i.e., resulting in the creation of fewer or more layers. During subsequent data analysis, layers from different trenches may be combined into *conceptual layers*, aiming at the recognition of entire 3D regions of the excavation space. This characterization is highly subjective and is again attributed to specific persons (archaeologists).

3.2 Finds and Recording Position

Surveying the finds in the following list, we identify the types of spatial objects that exist in this application context.

- (i) *samples* – samples that are received in connection with other finds for further study, e.g., for carbon dating (C14). Rough positioning suffices to record the origin of a sample.
- (ii) *postholes* (with wood) – specific type of find that exists only for lake settlements as it is the case in Dispilio. These holes are traces from the wooden posts supporting the houses of this specific settlement. Postholes often still contain wood pieces. Postholes are circular in shape.
- (iii) *scattered find* – items that were found spread over a larger area but as a whole constitute a single find. Its position cannot easily be recorded by means of a point location. Examples include pottery sherds, bone fragments.
- (iv) *small finds* – entire items that were found at a specific position (bone fragments, tools (iron, stone)).
- (v) *large finds* – objects that are of larger extent, e.g., hearths or walls.

As was evident from the excavation process, different methods exhibiting various degree of accuracy were used to record the position of finds. The following list surveys categories of positions, related degrees of certainty and measurement techniques.

- **Absolute positions** (small finds, postholes): For many finds the position is given in terms of a point coordinate. A point coordinate is measured using strings and/or wooden rules with respect to the sides of a trench.
- **Approximate positions** (scattered find, large find, samples): The positions of finds are recorded in terms of an approximate region within the trench they were found in. As we will see later on, such positions will either be recorded by rough direction estimates with respect to the center of a trench (samples) or by using sketches as in the case of scattered finds and large finds.
- **Unknown positions**: Given that positional information is missing from finds, it can be in the worst case approximated by the spatial extent of the section it was found in, i.e., with equal probability the position of the find is anywhere inside the given trench.
- **Depth**: Generally, the depth information is known since it refers to the depth of the excavation during a specific day. Should the depth not have been recorded, it can be approximated with the maximum top and minimum bottom depth of the section during this day of excavation. Should the coordinates of the find be known more precisely, the respective depths of the closest corner points can be used.

The uncertainty of positions in the excavational data is not provisional, i.e., it is not expected that that more accurate values will eventually replace approximate ones, or to be corrected. Thus, two provisions to handle unknown values are as follows.

- **Diaries:** The bottom dimensions of an excavation in a specific trench after a specific day should be identical to the top of next day's excavation. Thus should any of the depths be missing in a diary entry, the system should support a substitution of values based on "adjacent" diaries (chronological order).
- **Missing position:** The not-known position of a find can be substituted by the section it was found. This effectively leads to a fuzzy position whose probable position can be anywhere within the respective section.

For the final system prototype the provisions for dealing with uncertainty are to consider spatial uncertainty (a) in the underlying data model and (b) in the visualization of uncertain positions in general and in query results.

Approximate positions

Two types of approximate positions have to be recorded in the system, (i) cardinal directions and (ii) based on sketches.

Cardinal directions are a means for archaeologists to record the approximate position of a find with respect to the center of a trench. Fig. 2 gives the nine directions as used to denote the location of a find and their spatial coverage with respect to a trench by means of a regular grid composed of 3x3 tiles. To smoothen the discrimination effect between these categories, a 10% overlap between the cells is permitted (cf. the gray-shaded area in Fig. 2 showing the spatial extent of the center (C) cell).

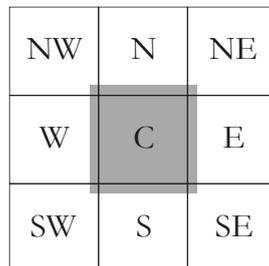


Fig. 2. Cardinal directions and their spatial extent

Storing these positions, we resort to spatial data types, i.e., should the user specify the position of a find as "Center", a polygon with the extent of the gray square in Fig. 2 will be stored.

Sketches are being used if (i) positioning technology is not available, (ii) in cases in which approximate positions suffice and (iii) the position of aerial features are recorded. Given that in the excavational process the rather small trench (5x5m extent) is used as a point of reference, such sketches generally produce accurate positions. In our approach for using sketches, an approximate position is recorded by means of a grid subdividing a trench into 10x10 cells. Each cell consequently has a side length of 50cm. As we will see later on, this approach is supported by a respective tool in the system in which cells covering the extent of the find are colored by mouse clicks (cf. Fig. 3).

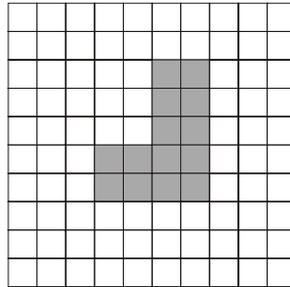


Fig. 3. Positional sketch

3.3 Trenches

Trenches as a means to structure the excavation space represent a simple geographic reference system for recording the position of finds.

A survey map of the trenches in Dispilio is shown in Fig. 4. Fig. 5 gives an impression of a trench that is excavated. Trenches that have been or were excavated are numbered. The grid is extended as needed, e.g., initially and starting in 1992 only the Eastern trenches were excavated. Later on the excavation was extended to the Western trenches and in recent years the Southern trenches were added.

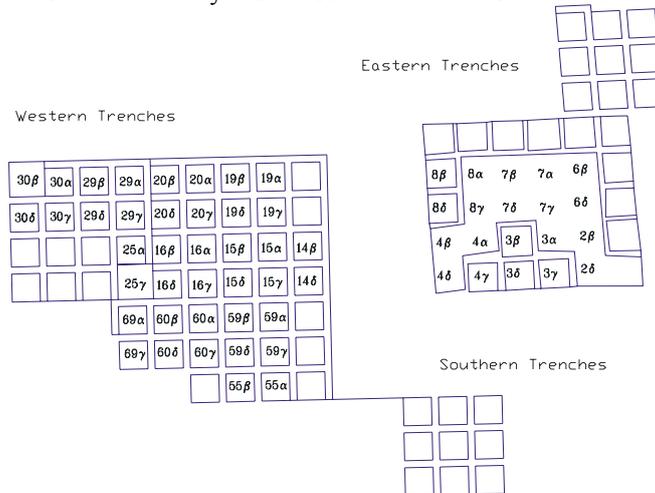


Fig. 4. Excavational grid and survey map for the Dispilio excavation

An initial problem was that no surveying techniques were used to delimit the trenches as is evident for, e.g., trenches 4γ, etc. Consequently, for this early data no exact survey map is available and these trenches are currently being surveyed again.



Fig. 5. Trench example²

Measuring positions

Upon discovering a find, it is important to immediately determine its positions. In the simplest and most typical case that of a small find, its point position is determined in terms of a perpendicular distance to two respective sides of the trench. Although the specific excavation guidelines state that the measurement should be taken with respect to the South-West corner, measurements taken with respect to any corner can be found in the collected data. This corner is recorded in the excavation diary, e.g., SW as in the example of Fig. 6(a) showing the typical square 5x5m trench and a small find with its point position.

The challenge we faced in developing an excavation system that should visualize the data for subsequent analysis was to translate all trench-relative coordinates to a global reference system such as WGS84. The task at hand was to develop an algorithm to translate any recorded position to a global reference system based on the existing survey map as shown in Fig. 4.

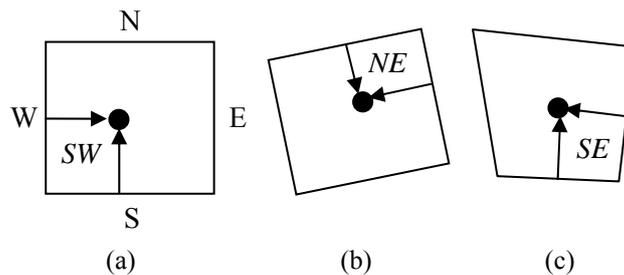


Fig. 6. Measuring position

Besides the reference point for the positional measurement, an additional problem for such a conversion is the position and shape of the trench itself. Although most trenches have a 5x5m square shape, they are rotated with respect to cardinal

² Picture © Dispilio excavations 2001.

directions. Moreover, trenches surveyed in the early days of the excavation are quadrilaterals rather than squares, e.g., Eastern trenches in Fig. 4. Fig. 6(b) and (c) give examples of positional measurements taken in rotated and non-square trenches.

Based on the above analysis, we developed an algorithm that takes (i) the trench-relative position, (ii) the direction the measurement was taken from and (iii) the trench geometry from a survey map (WGS84 referenced) as input parameters and outputs the position with respect to the WGS84 global reference system. The algorithm was implemented as PL/pgSQL [8] script for the PostgreSQL database used in the system prototype.

As a policy to record all original data and also for consistency reasons, both positions are kept in the database, the original trench-related and the translated position.

3.4 Data Model

Describing the basic entities and their spatial properties, Fig. 7 and Fig. 8 show data model excerpts for the various types of finds and diary. In the diagrams, the cardinality of a relationship is identified by the following symbols: many - \supsetneq , one - \dashv , the “o” symbol characterizes an optional relationship. Primary and foreign keys are marked with (PK) and (FK), respectively.

All finds have attributes *Position* and *Depth*. Depending on the type of find, *Position* will be implemented as a point (SMALL_FIND, POLEHOLE) or as a polygon (SAMPLE, LARGE_FIND, SCATTERED_FIND).

POLEHOLE has additional attributes *Height*, to record the depth of the whole representing the actual height of a wooden pole that created it, and *Diameter*, to record the size of the hole. The position of a SAMPLE is typically recorded in terms of cardinal directions and, using the model of Section 3.2, translated to a polygon. The positions of LARGE_FIND and SCATTERED_FIND are recorded by means of a grid (and respective tool) as illustrated in Section 3.2 and stored as polygons.

Since positional measurements are relative to a trench, i.e., with respect to one of the four corners of a trench, for SMALL_FIND and POLEHOLE the corner which was chosen in each case is recorded in POSITIONAL_REF_TYPE.

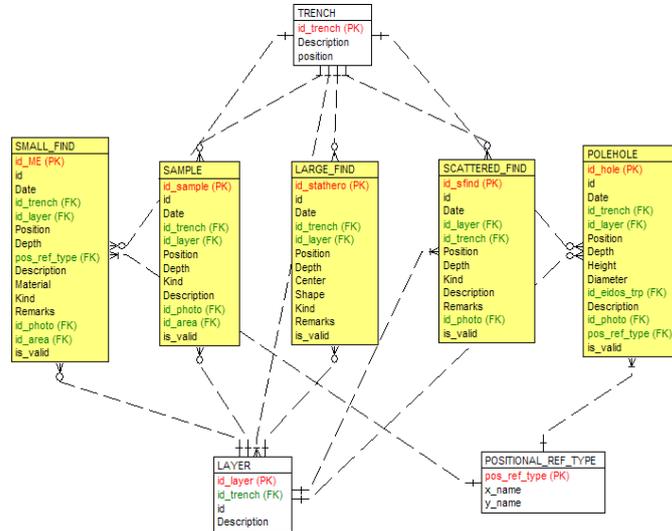


Fig. 7. Conceptual data model: finds and position

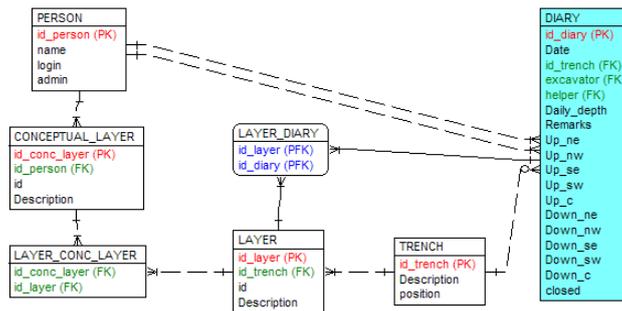


Fig. 8. Conceptual data model: diary and position

The fact that all finds are related to a specific trench is recorded by the respective relationships. For a TRENCH, its *Position* is stored as a polygon (cf. Section 3.3).

Excavation depth is more than a spatial measure for archaeologists. As a qualitative measure it represents different historic periods. To discover this fact in the excavational process, each layers are recorded with respect to diary entries as the dig progresses (DIARY, LAYER_DIARY, LAYER, TRENCH relationships). During a subsequent modeling phase layers across trenches are combined into conceptual layers (cf. also Section 4.3). Given that this modeling of layers is a highly subjective process (opinions differ among archaeologists), conceptual layers are recorded on an individual basis (PERSON, CONCEPTUAL_LAYER relationship).

Finally, although most information related to the excavation progress is kept in the diary, i.e., for each day excavating a specific trench, exactly one diary entry is created, the only spatial property recorded is the depth of the four corner points of a trench and the center are recorded at the beginning and at the end of the day.

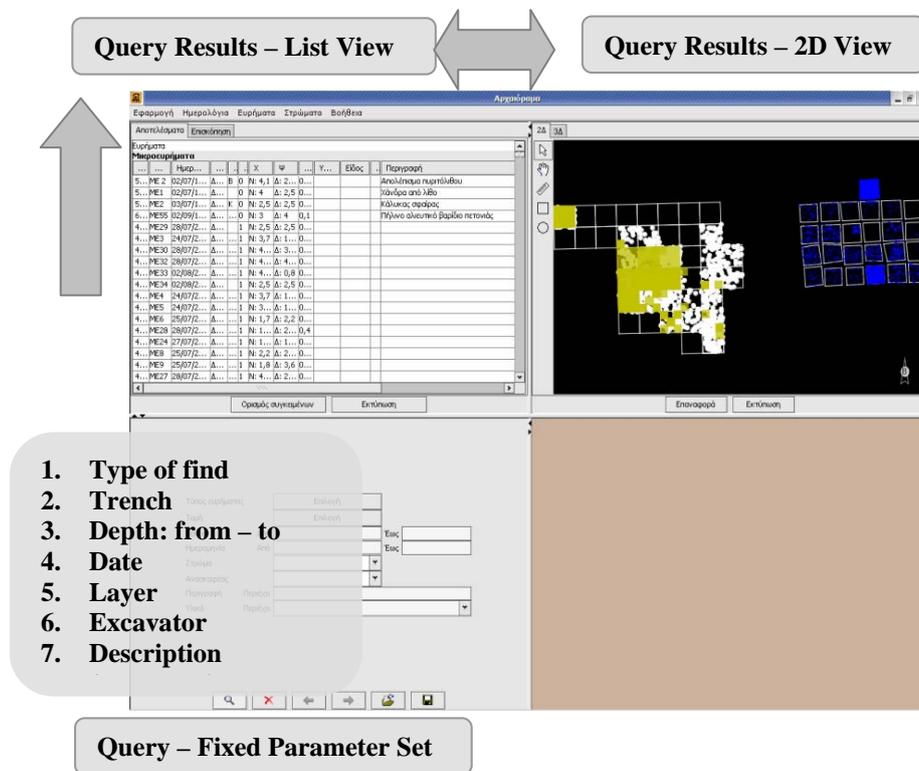


Fig. 9. Querying overview

4 Visualization and Querying

The basic means to explore the data stored in Archaiorama is by using simple parameter based queries. As shown in Fig. 9, by specifying values for up to seven parameters, all finds in the database are queried and the result set is presented (i) in a list view, (ii) as 2D representation and (iii) 3D representation of the positions of the finds in the result set.

Using the position one can essentially compose a collective view of the data, termed fields. An example would be to show all small finds for the whole excavation or selected trenches. Here based on a positional argument, data is retrieved. The system incorporates full spatial query functionality allowing for arbitrary spatial selection in two dimensions. Fig. 9 shows the query interface, where

- in the bottom-left corner, the various parameters are specified (type of find, trench, depth, date, layer, excavator in charge, description, material),
- in the upper left corner the results are presented as lists and

- in the upper-right corner, the query results are presented in a 2D map-style representation and in 3D.

To quickly distinguish the various types of finds in the 2D and 3D representation, the objects are visualized in different colors, i.e., white – small finds, blue – postholes, yellow – samples, light green – large finds, red – pots, orange – ceramics (sherds), gray – stone tools, purple – construction tools, dark green – bones, light blue – shells.

4.1 2D Visualization and Uncertainty

The accuracy of a recorded position can be deduced from the shape of the position. Fig. 10 gives an example of a parametric query result in which discovered items are shown that have (1) approximate positions (drawn using the positional grid editor), (2) unknown positions (only the trench is known) and (3) known, precise positions (x,y co-ordinates).

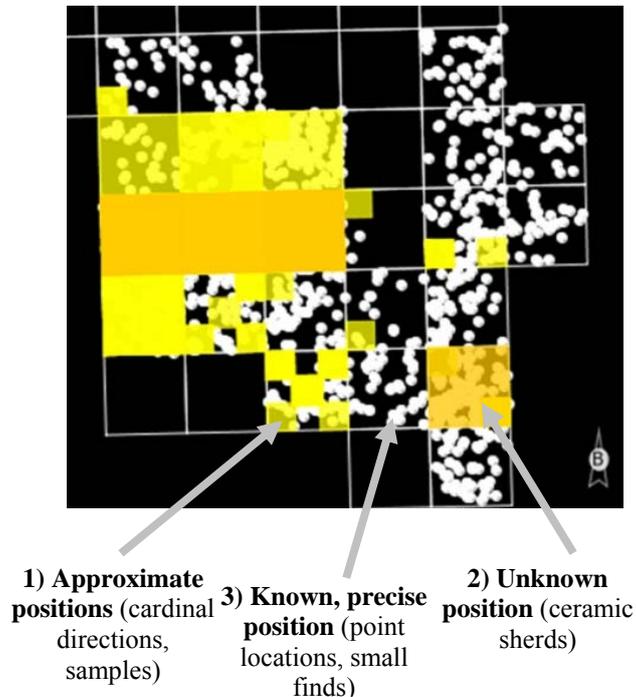


Fig. 10. Examples of positional certainty

Fig. 11 shows how uncertainty is shown with respect to query results. Given an approximate position, the percentage by which the position of the respective item is contained in the query window is shown as a mouse roll-over window. This percentage can be perceived (i) in the case of an approximate position as the probability by which the item would be located in the query window or (ii) in case of aerial features, the percentage of the entire item located in the query window.

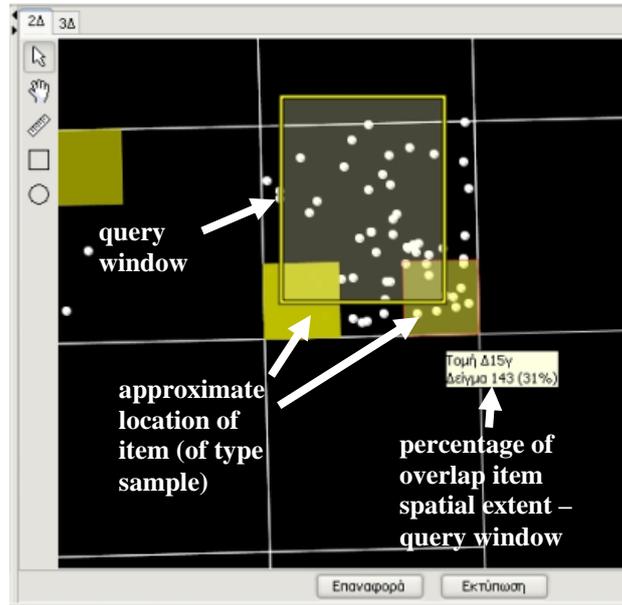


Fig. 11. Range queries and approximate query results

The current prototype implementation of Archaiorama does not distinguish between these two types of objects. This has to be decided by the user in connection with the type of discovered item.

4.2 3D Visualization

Visualizing two-dimensional data in three-dimensions (plain + depth) is fully available in the system. The result of an example query is visualized in Fig. 12. The

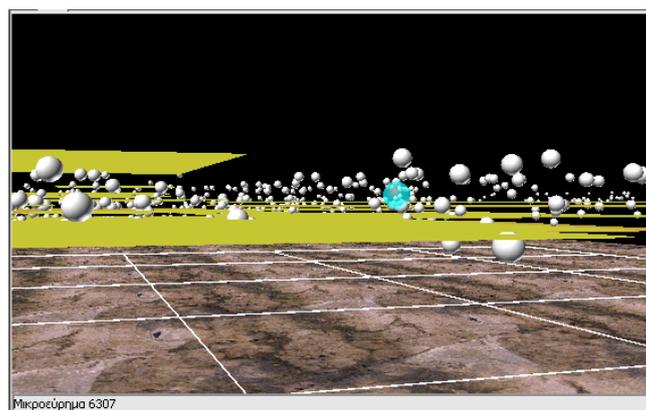


Fig. 12. 3D Visualization with object selection (blue dot)

only supported query functionality from within this representation however is object identification, i.e., using the mouse one can select any object and get its related object identifier in the legend.

4.3 Special Visualization Aspects – Layer View

Layers are recorded in the header of each diary and relate to the trench the respective diary refers to. Layers can be grouped in conceptual layers, which essentially can be used to semantically model the excavation space, i.e., relating different historic periods to portions of the 3D excavation space. This grouping (clustering) is purely subjective and different for each user.

Management of layers

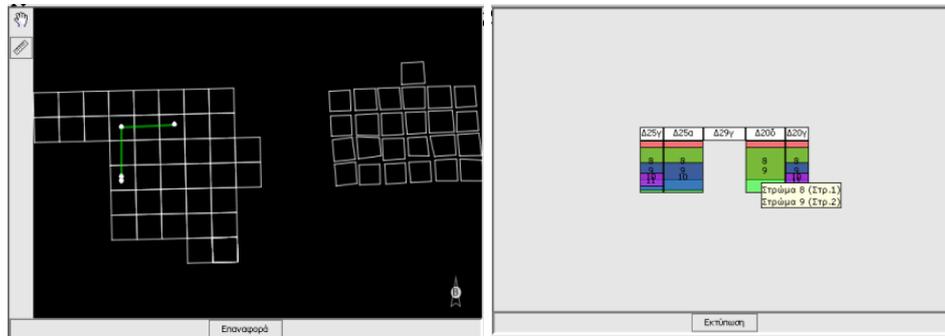
Archaiorama provides a tool for creating conceptual layers and to visualize the layer information recorded on a trench basis in the diary entries (cf. Fig. 13). Conceptual layers are created by concurrently selecting layers from several trenches and assigning a global identifier and description. Layers that belong to a conceptual layer are indicated in pink as opposed to white being used for ungrouped layers. Recorded conceptual layer information, number and description, is displayed in a tool-tip window. This information can be changed at any time and layers can be added and/or deleted from a conceptual layer.

The screenshot shows a grid with columns labeled Δ02α, Δ02β, Δ02γ, Δ02δ, Δ03α, Δ03β, Δ03γ, Δ03δ, Δ04α, Δ04β, Δ04γ, Δ04δ, Δ0εβ, and Δ0. The grid contains numerical data representing layers across different trenches. A tooltip window is open over the cell at row 7, column Δ03β, displaying the text "(Στρώμα 2: Δεύτερο στρώμα)". Below the grid, there is a control panel with a dropdown menu for "Αριθμός" (Number) set to 2, a text input for "Περιγραφή" (Description) containing "Δεύτερο στρώμα", and two buttons labeled "Νέο" (New) and "Ορισμός" (Define).

Fig. 13. Creation of conceptual layers

Visualizing conceptual layers

Conceptual layer information can be visualized by taking arbitrary cross-sections of the excavation space. This is supported by a query tool that draws a polyline on top of a map showing the trenches of the excavation space (cf. Fig. 14(a)). Based on this cross-section, a separate window displays the query result, i.e., the (conceptual) layers



(a) layer query: excavation map (trenches) and cross-section

(b) query result: colored-coded excavation layers

Fig. 14 Querying layer information

for the selected trenches in the sequence the cross-section was drawn (cf. Fig. 14(b)). Again, tool-tip windows are used to display layer information. Different conceptual layers are identified using different colors.

5 Conclusions and Future Work

The creation of an excavation documentation system is a challenging task given the heterogeneity of requirements to such a system. This work describes the spatial data management aspects of the prototype system Archaiorama that is currently being used for excavation documentation in Dispilio, Greece. Surveying the types of excavation data, their (spatial) properties, and the particularities of the excavation space lead to the definition a respective data model. Data analysis in Archaiorama is supported through spatial query capabilities in connection with 2D and 3D visualization options. Finally, a brief system overview should complement the specific descriptions in this paper.

The directions for future research are as follows. Spatial query capabilities as such are so far limited to range queries. An important aspect in data analysis however, would be to support data mining functions e.g., clustering. In the current data model, the excavation space is not modeled as 3D but rather 2D plus depth. For larger databases (currently only ~3000 finds and 700 diaries are kept in the database), such a shift could result in performance gains. Similarly, the current database system does not support three-dimensional data types. However, for several finds (large finds, poles) such a representation would be useful. Further issues of future work relate to performance measurements for concurrent use also through network connections, integration of wireless clients, and providing interfaces for GPS receivers and other positioning tools.

Acknowledgements

This research is supported by the “Arxaiorama” project, funded by the Greek General Secretariat of Research and Technology.

References

- [1] ArcTron. ArchaeoDATA. Product web page, <http://www.arctron.com/Software/ArcheoDATA/>, current as of June 2006.
- [2] Dekoli, M. and Hadzilacos Th. A GIS- and hypertext-based system for excavation documentation. 25th International Conference of Computer Applications and Quantitative Methods in Archaeology (CAA), Birmingham, 1997.
- [3] Dekoli, M. and Hadzilacos, Th., 2002. The utilization of computer technology in archaeology: Dispilio 1998-2002. Proceedings from the Theoharis conference, Thessaloniki, 2002.
- [4] Delis V., Hadzilacos Th., Tryfona N., An Introduction To Layer Algebra, 6th International Symposium on Spatial Data Handling (SDH), Edinburgh UK, 1994.
- [5] Dispilio Excavations. Project homepage, <http://web.auth.gr/dispilio>, current as of June 2006.
- [6] Hadzilacos, Th. and Stoumbou, P. M. Conceptual Data Modelling for Prehistoric Excavation Documentation. 23th CAA conference, Analecta Praehistorica Leidensia no. 28, 1996.
- [7] ICOM/CIDOC Documentation Standards Group. CIDOC Conceptual Reference Model. Available at <http://cidoc.ics.forth.gr/>, 2005.
- [8] PL/pgSQL - SQL Procedural Language, PostgreSQL Manual, <http://www.postgresql.org/docs/8.1/static/plpgsql.html>, current as of June 2006.
- [9] PostGIS. Project homepage <http://www.postgis.org/>, current as of June 2006.
- [10] PostgreSQL database. Product homepage <http://www.postgresql.org/>, current as of June 2006.
- [11] Singularch. A data collection system for archaeological excavations. Product web page, <http://singularch.de/software.htm>, current as of June 2006.