Deconstructing the Product into Theory

Cesar A. González-Perez

Laboratorio de Arqueología y Formas Culturales
Instituto de Investigaciones Tecnológicas
Universidad de Santiago de Compostela
phcgon@usc.es

Abstract. Information systems used in different fields must be based upon and reflect the theoretical and methodological underpinnings of such fields. This paper shows how the SIA+ (Archaeological Information System, enhanced version) has been designed and built around theoretical foundation regarding information structure, space, time and subjectivity management; and using methodological bases such concerning information availability and work dynamics.
Keywords: information systems, SIA+, CVS, space management, subjectivity management, time management, information availability, work dynamics.

1 Introduction

This paper presents the SIA+ (Archaeological Information System, enhanced version), a software product developed by the Laboratory of Archaeology & Cultural Forms (University of Santiago de Compostela, Spain) for internal use. This laboratory employs over forty people, five of them working in information technologies. The LACF specialises in applied research and services for third parties, including assessment and correction of archaeological impact, social enhancement of cultural heritage and information systems for cultural resources management. The LAFC also keeps a general basic line of research on landscape archaeology.

Basic research performed in the field of information technologies has focused on software development support (González 1998a, 2002), representation systems (González 2002, 1997a), and time & space management (González 1999a). Applied research conducted on these bases is rendering information structure and work dynamics management systems, and has already provided results such as the CVS technology (González 1999a). Built on top of these developments, the SIA+ uses a bit of all of them.

2 The Product

The SIA+ covers a number of different information areas such as excavations, stratigraphy management, archaeological impact assessment, finds management, report generation, bibliography and staff & projects management. It is structured around 37 different forms and 13 searching, reporting, import/export and collaboration integrated tools. Currently, it stores information about over 3,500 archaeological sites, 49,000 finds, 17,100 photographs, 6,100 archaeological & heritage assessments, and 230 projects. These quantities may not be very large for a region- or countrywide system, but they are significant for a system designed and built to support the daily work of a single research group. Also, notice that the Galician archaeological record is characterised by a low average number of finds per excavation, and that the vast majority of the projects within the LAFC involve field surveys but no excavations at all.

The system's user interface reflects well some of its characteristics. Information is grouped into folders so the user can easily access the necessary forms. Forms are the main data entry and browsing mechanism, and make extensive use of conventional fields and controls. Also, search tools allow the user to perform customised searches and send the results to text documents or to object collections, which may be used later to perform block operations on groups of objects. Integrated mapping allows the user to locate entities on a map and measure distances and areas. Finally, task lists provide the main way for inter-user coordination.

For those interested in the technical details, the SIA+ has been built using Microsoft Visual Basic, and it stores information into a Microsoft Jet database. The resources needed to complete the project included one person during seven months, although a number of modifications have been done afterwards, making an accurate estimation of effort very difficult. The typical use of the system is of ten simultaneous users on average with twenty-six at peek times. The server computer in which the database is hosted is a dual Pentium Pro at 180 MHz with 128 MB of RAM, and the client workstations from which the SIA+ is used range from Pentium 90 up to Pentium II 450 with 32 to 128 MB of RAM. The operating system is Windows NT 4 in all cases.
The SIA+ user's manual has been published as González 1997b.

3 Theoretical Foundations

The SIA+ has been built on top of some theoretical foundations such as information structure, space management (and the related CVS technology), and time & subjectivity management.

3.1 Information Structure

Knowing the structure of the information involved in a computer system is crucial to its success, and some previous work was taken into account when modelling the archaeological concepts and relationships that were defined by the future users of the SIA+. Specifically, class models with over 280 classes had been developed two years before (González and Bóveda 1998), and a glossary with over 420 terms had been built. A fully object-oriented approach to modelling was chosen, using a modified version of the Fusion method (Coleman *et al.* 1994; Malan *et al.* 1996; González 2002), and Microsoft Visual Basic was used to implement the program. Visual Basic is a generally good environment, but its object-orientedness is somehow special; for example, no inheritance mechanism is provided, so specialisation hierarchies must be implemented by aggregation of interfaces, which we have found a more clumsy and time-consuming technique. Finally, the Microsoft Jet relational database engine was used to store the data. The process of changing paradigms from fully object-oriented to pure relational as the development goes on is usually known as impedance mismatch, and made our work much more difficult than it would have been in a pure object-oriented development environment.

The class models mentioned above nearly covered all the needs of the product and, in fact, were simplified in order to build the SIA+. Some conventional tricks were used to map classes and relationships into relational tables, such as mapping every specialisation to a 'parent' table corresponding to the superclass and one more 'child' table corresponding to each of the subclasses. A unique key mechanism was used to guarantee object identity.

At the user interface level, each class is represented by a form containing specific fields and lists, which can be easily mapped by the user to the class' attributes and relationships. Relationships can be established through a drag and drop mechanism, and broken by clicking a button.

3.2 Space Management and CVS

After working with archaeologists for some years, it became visible that a friendly and powerful mapping integrated system was needed. First, it should provide a 'continuous surface' approach to representing the earth's surface, avoiding the need to change map sheets after hitting a sheet border. Also, the system should be able to work at different cartographic scales, approximately ranging from 1:2,000,000 up to 1:10. The user should be also relieved of the work of manually changing sheets depending on the desired scale. Finally, the system should be fast enough to be useful. All these factors have been discussed in González 1998b.

Taking all these needs together, we developed the CVS technology, which stands for "segmented vectorial cartography". It is comprised of a file format, an access library, and some additional tools (González 1999a). CVS offers horizontal cartographic integration, as it can retrieve any segment of the available area to be displayed or processed via a two-dimensional spatial indexing mechanism that divides the available data into discrete sectors, and do it fast. Also, it offers dynamic zooming, which automatically provides more or less cartographic detail depending on the working scale.

Basically, the CVS file format stores a thematic layer on each file, including several levels of detail and the two-dimensional index. CVS files are highly compacted binary files with a complex internal structure, so the user will not want to edit them manually. A DXF-to-CVS converter has been developed, which can take a DXF file as input and create a CVS file as output, allowing the user to specify customised mappings between DXF layers and CVS levels of detail, as well as to set appropriate indexing parameters. Currently, the conversion process takes a long time and needs huge amounts of disk space, but we are working to reduce these requirements.

The CVS access library 'knows' the CVS file format and provides an ActiveX interface to CVS data to any application that needs it. It exports the classes Layer, Levels, Level, Sectors and Sector, corresponding to a whole CVS file, its collection of levels of detail, each level, each level's collection of sectors, and each single sector, respectively. Also, a CVS editor has been developed to let users examine and edit the contents of CVS files.

3.3 Time and Subjectivity Management

Few information systems in the market offer good time management, and even fewer can manage subjectivity. Both points are vital for a system designed to work in a context of humanities, and especially in archaeology, where absolute truth can be seldom recognised, and the flow of time and personal judgements are part of the usual processes for knowledge construction. From a practical perspective, the future users of the SIA+ needed to cope with the changes that objects experiment over time, and support simultaneously an arbitrary number of different subjective viewpoints (González 2002).

The solutions implemented include object versioning, the use of timestamps in every versionable object, and time-sensitive searches. Object versioning, a difficult issue, has been implemented in a very straightforward way, trading functionality for ease of use and development time. Any object of the valorative classes supported by the SIA+ (archaeological assessments, finds assessments, heritage assessments, heritage situation, impact assessments and corrective measures) can be versioned by any user by clicking on a button. The system saves the newly entered information together with the user's identification and a timestamp. At any moment, any user can request a list of an object's versions, making the system show the author and time-

stamp for each of them. Of course, the contents of any of them can be easily retrieved.
Searches are time-sensitive so the user making the search can chose between a historical search and a point-in-time search. The first type would show all the versions of the objects resulting from the search criteria up to a specific date, while the second type would show the last (relatively 'current') version of each object resulting from the same criteria to a specific date. These specific dates can also be specified by the user, thus allowing 'arbitrary time' searches that show the state of the information at any past point in time.

4 Methodological Foundations

Apart from the theoretical foundations, the SIA+ has been built on top of some methodological bases such as information availability and work dynamics.

4.1 Information Availability

The availability of information in any system must follow a good signal-to-noise ratio, avoiding situations in which the needed information does exist in the system but is very difficult to locate or use. Also, good search and navigational capabilities are needed in order to facilitate the retrieval of related information.

The theoretical principle of contextual interaction (González 1999b) was used to model the user interface, so the active user can restrict the visible information to that related to the current project he is working on by choosing that project from a master list. The rest of the information (not concerning the selected project) is totally hidden to the user, so the 'noise' decreases while the 'signal' remains the same.

Several search tools were developed into the system, with different orientations and varying degrees of complexity. A full-text search tool is available, so any word or sentence can be sought for in the whole database. The system does not use inverted dictionaries or other ways of optimising full-text search, so this operation can take a long time. A multi-threaded mechanism was used to prevent searches from blocking the system. Also, a 'favourites' list is available so every user can bookmark his most visited objects in the database for fast and direct retrieval afterwards. As a third tool, a filter can be set on any class, so only the objects matching the filter criteria are shown. These criteria can be freely specified by the user through an easy-to-use graphical interface. The resulting set of objects can be 'saved' as an object set, and come set operations are available to unite, intersect and subtract on set from another.

Finally, a highly customisable reporting engine has been integrated into the SIA+, allowing the users to specify a root object from which the search engine traverses relationships, collecting information on every object visited and assembling a formatted Microsoft Word document with the results. Up to sixteen different options can be specified, ranging from time-sensitiveness to what classes and attributes are to be recorded in the final document, and including parameters for the breadth and depth of the traversal search. The resulting document can be printed, saved or edited in any conventional way.

An additional way of improving information availability was achieved by the use of hyperlinks. Any field or list in a form representing a relationship has an associated button that can be clicked to navigate to the related object. Also, any object code appearing in any text body can be used as a hyperlink by clicking on it. This results in a very fast way to visually navigate the relationships among the objects in the database.

Finally, all the preferences and settings of each user, including favourite bookmarks and contextual interaction parameters are stored into the user's Windows NT roaming profile, so they travel with the user from computer to computer as needed.

4.2 Work Dynamics

A very early demand from the SIA+ users referred to the need of not only a structural system, but also a functional one. Conventional databases present the user a (more or less appropriate) model of the application domain, which he can use to support the everyday work. A functional system would offer this structural support and, in addition, would know, deploy, track and measure the routine processes that take place during daily work. A key point to achieve this is, of course, good inter-user communication mechanisms.

Some work has been done on workflow modelling, and a workflow engine is currently under construction. The solutions implemented, however, include intranet-based specifications and multi-user task lists. Some support of workflow tracking is planned for the next version of the SIA+.

Intranet-based specifications can help achieve a good quality of information, by providing rules and guidelines that every user must follow when describing objects or entering information into the system. The intranet in the LAFC contains specifications on how to describe different kinds of objects, an extensive glossary, and suggestions & problems report forms. It can be accessed from a conventional web browser or directly from the SIA+, thus allowing a user to retrieve procedural information without leaving his working environment.

Multi-user tasks list, on the other hand, provide the major way for users to co-operate and communicate. Any user can post pending tasks to another user, including himself. A pending-tasks red indicator comes up when the currently active user has pending tasks, so he can review the task list and mark any of them as completed. Deadlines, notification thresholds and priorities are user-definable.

5 Summary

The SIA+ is a big system that has been used by many people during a long time in an integrated work environment, so we believe that the theoretical and methodological foundations that support it may be right. Of course, much work is in progress, and the SIA+ will be soon upgraded both functionally and technologically.

References

Coleman, D., Arnold, P., Bodoff, S., Dollin, C., Gilchrist, H., Hayes, F. and Jeremaes, P. 1994. *Object-Oriented Development: The Fusion Method.* Englewood Cliffs (NJ). Prentice-Hall.

González Pérez, C.A. 1997a. Beyond Literacy: The Digital Management and Production of Archaeological Record Archives. In: Criado Boado, F. and Parcero Oubiña, C. (eds), *TAPA (Trabajos en Arqueología del Paisaje) 2.* Santiago de Compostela (Spain). LAFC, IIT, USC.

González Pérez, C.A. 1997b. *SIA+: Manual del Usuario.* CAPA (Criterios y Convenciones en Arqueología del Paisaje) 3. Santiago de Compostela (Spain). LAFC, IIT, USC.

González Pérez, C.A. 1998a. Submarines, Classification, and the Fusion Method. *Fusion Newsletter* 6.1 (April 1998). Palo Alto (CA). Hewlett-Packard Company.

González Pérez, C.A. 1998b. *Modelos de Paisaje para Arqueología: GIS y Teoría.* gamdds (Grupo de Amigos do Museu D. Diogo de Sousa). Braga (Portugal). Museu D. Diogo de Sousa.

González Pérez, C.A. 1999a. The CVS Data Format. *Dr Dobb's Journal* May 99. Miller-Freeman.

González Pérez, C.A. 1999b. Proactive User Interfaces for Cultural Resources Management Systems. In Barceló, I Briz and A. Vila (eds), *New Techniques for Old Times. CAA98 Computer Applications and Quantitative Methods in Archaeology* (Proceedings of the 26th Conference CAA 98). BAR International Series 757. Oxford (UK). Archaeopress: 297-298.

González Pérez, C.A. 2002. *Sistemas de Información para Recursos Culturales: Teoría, Metodología y Tecnologías.* British Archaeological Reports (BAR) series. Oxford (UK). Archaeopress.

González Pérez, C.A. and Bóveda López, M.M. 1998. Un Modelo de Clases para el Registro Arqueológico. *Boletín del Instituto Andaluz del Patrimonio Histórico* 22. Sevilla (Spain).

Malan, R., Letsinger, R. and Coleman, D. 1996. *Object-Oriented Development at Work: Fusion in the Real World.* Upper Saddle River (NJ). Prentice-Hall.