Information Systems for Cultural Resources Management: Theory, Methodology and Technologies

César A. González Pérez

Santiago de Compostela, December 1999

Table of Contents

Preface

This work aims to develop a technology that will allow the design and construction of information systems applied to cultural resources management. Including terms such as technology, design, information systems, management and cultural resources in the same sentence, and making this the main goal of a thesis implies approaching the task from a multidisciplinar, self-critical and humble viewpoint. Therefore, this work dives deep into topics such as cultural resources management or theoretical computing, gets close to some others such as human-computer interaction, and merely scratches the surface of others such as philosophy and software development. At the same time, and partly as a consequence, this work often regrets deep research in benefit of concept integration, or searching for the primary source of an idea to deal with its new applications in new fields. We must not consider this as a lack of merit but as an alternative way of working, which prefers widening instead of diving.

This work is structured as follows:

- Chapter I, *Introduction*, presents the objectives of the thesis and reviews the current state of the art.
- Chapter II, *Theoretical and Methodological Proposal for the Application Domain*, describes our conceptual scheme for the practice of cultural resources management.
- Chapter III, *Theoretical and Methodological Proposal for the Computing Domain*, explains our ideas on how to tackle information systems for cultural resources management.
- Chapter IV, *Information Models*, presents the models obtained from the contents of chapter II by applying the constructions developed in chapter III.
- Chapter V, *Technological Proposal*, deals with which technologies can be used to implement the concepts developed so far. Some of these technologies alrady exist so how to use them is described; some others don't, so research guidelines are given.
- Chapter VI, *The Archaeological Information System*, describes a specific information system developed by the author of this work, and uses it to verify the positions posed in previous chapters.
- Chapter VII, *Consequences*, presents the consequences of this work, centering on both basic research and technological needs towards a near future.

Finally, a note on document conventions: “quoted” words or expressions refer to the term they express; *italicized* words or expressions refer to the concept they want to mean.

I. Introduction

I.1. Objectives

The main objective of this work is *to develop a technology for the design and construction of information systems in cultural resources management*. We can define *technology* as a specific body of available knowledge, usually in the form of products or services, that allows putting into practice theoretical designs. This definition makes us to deal with theoretical, methodical, methodological and contextual components during the development of such technology:

- Theoretical, because every mature technology needs supporting foundations beyond its commercial success or degree of applicability.
- Methodical, that permit verifying the coherence and adequacy of the discourse against the proposed goals.
- Methodological, because they allow to apply the technology to problems to get solutions.
- Contextual, because we must consider the development and application environment of a technology to make it work.

The proposed technology serves to design and build information systems, that is, specialized tool sets that deal with information in a specific field or subject and, in particular, obtain *information* from *data*. The realm of applicability or *application domain* of the proposed technology is the management of cultural resources, and this comprises much more implications than it may seem. In fact, *the appropriate application of information systems to humanities* accounts for the major body of this work, due to both its volume and the amount of innovation we believe it brings, and we think it must be taken as an underlying and abstract objective of this thesis.

I.2. State of the Art

Humanities pose a new set of problems concerning the application of information technologies, such as subjectivity and temporality. Altough there exist many computer applications for humanities, even for cultural resources management, which have been more or less functional and successful, we must take into account those new issues fully and avoid a "shoehorn approach" that would try to make existing technologies fit new needs beyond adequate standards.

We have reviewed more than thirty software products specially designed for cultural resources management, focusing in four variables:

- System width, measured by the number of information ambits involved. Very "narrow" systems do not require solid teoretical foundations to stay coherent, but wide ones do.
- System depth, determined by the number of different abstraction and classification levels in the system. This variable is a good indicator of the cognitive complexity of the system.
- System size, measured by the amount of money invested, number of people involved, or other suitable metric.
- Degree of technological innovation, gauged by the occurrence of the following technologies during the product's lifetime: object-oriented databases, evolutionary project lifecycle, application-domain theoretical transfer, document and workflow management, object-oriented modeling techniques, client-server infrastructures and geographical information systems.

The results of our analysis state that:

- System width ranges from small to medium. Large projects have been approached by building several more-or-less interconnected smaller systems. This is not necessarily a bad choice, but we would have liked finding some cases of wide systems.
- System depth is nearly always minimal, involving information at a single abstraction or classification level.
- System size is usually medium or large, often involving dozens of people, long development schedules and high funding resources. There were also small systems designed and built by even a single person, but scarcely.

- The degree of technological innovation is very low. Traditional technologies such as relational databases and word processors were almost the only tools used to construct the reviewed information systems.

From the information exposed, we think that *information systems for cultural resources management are being developed by non-specialized people and with insufficient technological support, consequently resulting in projects that, independently from their success or failure, demand excessive amounts of time, people and money*.

To solve this problem, we believe that some effort must be made to achieve:

- A wide theoretical framework that allows to insert, at a middle abstraction level, the structure and function found in the daily work of cultural resources managers.
- A methodological body capable of supporting the processes of information systems design and construction.
- A number of technological developments to help the information technology practitioner in his daily work.

This thesis will try to advance in the three above-mentioned directions.

II. Theoretical and Methodological Proposal for the Application Domain

The term "application domain" usually means the subject to which information technologies are applied. In our case, it always refers to cultural resources management, so in this chapter we will explore a proposal to get close to cultural resources management from a computer spacialist's perspective.

II.1. The Way of Theorizing

Knowledge production is not independent of the models it uses. Therefore we think that a strong relationship can be established between the application domain and the information technologies (or "computing") domain, using the actual process of modeling as link. Figure II-1 shows this relationship and its direct consequences.

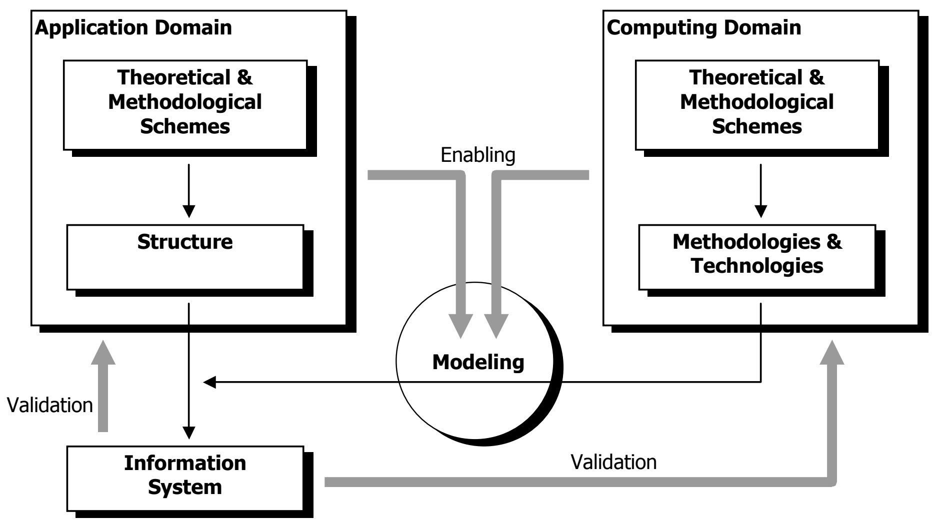

Figure II-1. Enabling-validation relationship between the application and computing domains. Theoretical and methodological schemes in both the application and computing domains, as well as the structure of the application domain and the methodologies and technologies used within the computing domain *enable* the construction of information systems through a modeling process. Simultaneously, the resulting information system *validates* the mentioned elements in both domains.

However, we must say before continuing that the "Information System" box in Figure II-1 does not show that every information system, in addition to reflecting the *structure* of the application domain on which it works, must provide some *functionality* that makes it a useful tool. This functionality must take place within the mentioned structure, giving rise to some relationhips between these two parts as shown in Figure II-2.

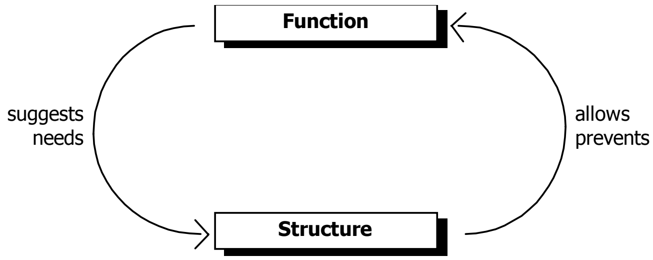

Figure II-2. Relationships between structure and function. Certain structures may allow or prevent certain functions; each function suggests or even needs certain structures.

Although the relations between the two parts are bi-directional, structure can be pre-existent and function can't: in general, *it is possible to build a generic structure with no function in mind, but it's impossible to design a function without referring to a supporting structure*. We will call this assertion Principle of the Pre-Existing Structure.

Figure II-1 and Figure II-2 will outline our journey from now on.

II.2. Theoretical Foundation

Starting at box "Theoretical & Methodological Scheme" inside "Application Domain" in Figure II-1, the thoretical foundation that we propose for cultural resources management is based in the *valorative chain* and its most direct consequence, the concept of *representation*.

Following Criado, three parts are involved in the formation of the archaeological record: a past society, some post-depositional processes, and a current social and institutional context. The archaeological record does not exist before perceiving it, but is created through a number of interpretative stages. Figure II-3 shows these stages and the processes that take place among them. merchandise

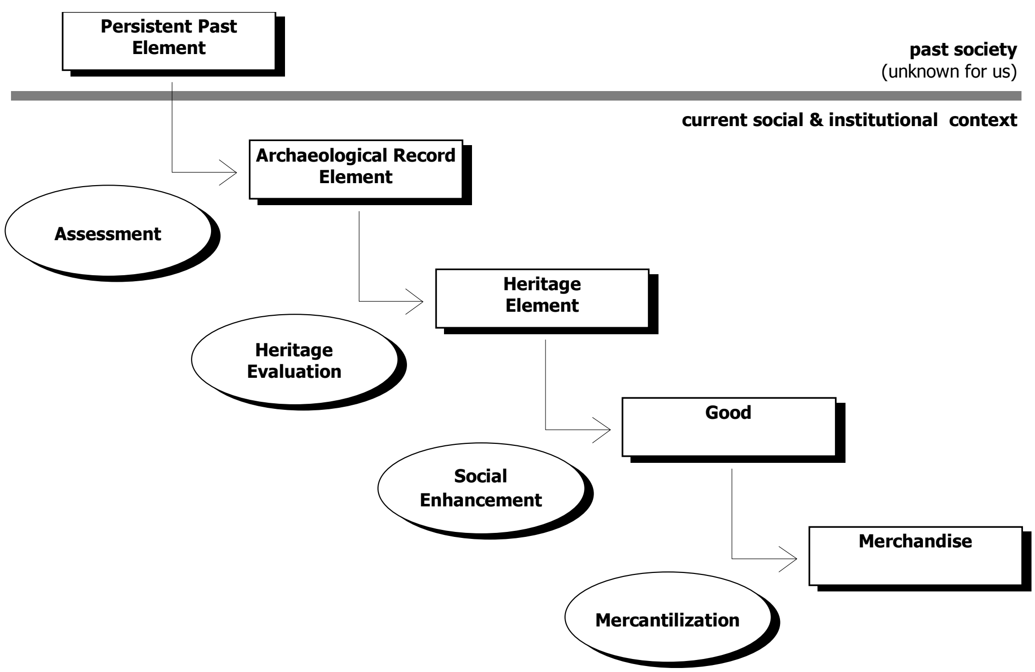

Figure II-3. The valorative chain. Ellipses represent interpretative processes, and boxes correspond to specific kinds of states in which a persistent past element can be found.

Each possible state of a persistent past element along the valorative chain is what we could colloquially call a "representation" of that element. We can detect two problems with this. First of all, what are the nature of things involved in representations?. Secondly, the term "representation" can mean something that exists and acts in the place of other thing, or the process of constructing that "something".

We can answer the first question introducing the concept of *entity*, that can be defined as *any portion of the observed reality with a degree of self-containment high enough as to possess its own identity*. We solve the ambiguity posed in second place by reserving the term "representation" for the process, and calling the product of this process a *referent*. The object on which representation processes work to produce a referent may be called a *substrate*. Notice that the very same object can act as referent and substrate, even at the same time; in fact, referent and substrate are not kinds of things themselves, but different roles that entities can play. Also, we must introduce the concepts of *conception* and *concept*. A conception is *an idea or perception that an individual has about an entity*, while a concept is *a conception shared by a sufficient amount of people as to assume a certain degree of intersubjectivity*.

Representation processes live within specific contexts, or semantic environments, that can be analyzed. FFF shows the parts involved in such representation contexts.

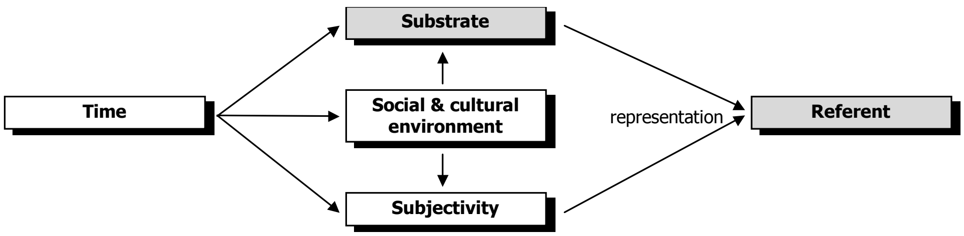

Figure II-4. Representation and its context. Arrows can be read as influence relationships. Substrate and referent are the main parts in the representation process. Time shapes substrates (destroying, changing and revealing entities), affects the social and cultural environment (by letting cultures change their values and knowledge assets), and also touches subjectivities (because individuals learn and change the way they act and think through their lives). The social and cultural environment selects substrates (giving birth to things) and powerfully moulds subjectivities. Finally, subjectivities create referents from substrates.

Finally, some words on structure and science. We can define structure as *the set of relationships that arise among some entities, plues the roles they play in these relationships*. Following this definition, we can describe referent validity in terms of similarity between its structure and the structure of its substrate. The closer they are, the more valid the referent is.

At the same time, scientific disciplines maintain two principles that are simultaneously goals: objectivity and reproducibility. Representation processes within sicentific realms cannot vary along the subjectivity or the temporality axes, letting referents change only accordingly to their substrates. Therefore, *representation processes in the context of scientific knowledge can be modeled by a subset of the concepts that are needed in humanities fields*. We will call this assertion Principle of the Scientific Simplicity.

II.3. Methodology

Our methodological approach to the practice of cultural resources management, corresponding to box "Theoretical & Methodological Scheme" inside "Application Domain" in Figure II-1, involves the following areas of activity:

- Cataloguing, location, and delimitation of cultural resources: you can only protect what you know.
- Evaluation, study and understanding of the historical and current meaning of resources: you can only know what you understand.
- Action and protection of cultural resources: you can only understand what you analyze.
- Restoration and enhancement of resources: you can only analyze what you evaluate.

II.4. The Observed Reality

The basic concepts that shape the structure of the application domain, corresponding to box "Structure" inside "Application Domain" in Figure II-1 are the following:

- Archaeological Record
- Archaeological Action
- Archaeological Find
- Archaeological Structure
- Stratigraphic Unit
- Site

- Archaeological Point
- Heritage Group
- Archaeological Assessment
- Heritage Evaluation
- Impact Assessment
- Corrective Measures

III. Theoretical and Methodological Proposal for the Computing Domain

This chapter explains our point of view on how the information technologies must be applied to the cultural resources management field from a theoretical and methodological perspective.

III.1. Theoretical Foundation

Referring to box "Theoretical & Methodological Scheme" inside "Computing Domain" in Figure II-1, we can say that an information system is a specialized set of tools designed to work within a specific field and obtain information from data. Figure III-1 shows how information systems relate to their environment.

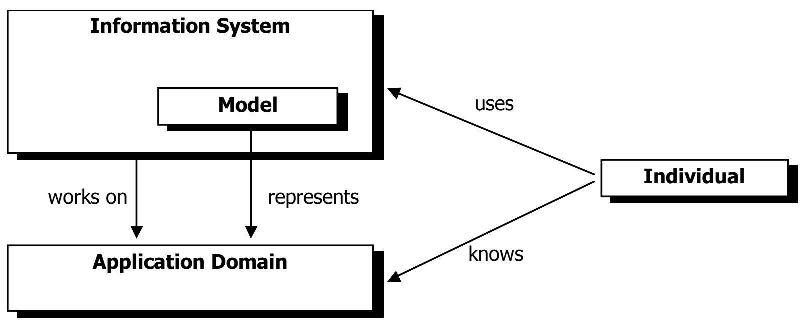

Figure III-1. Every information system must model the application domain on which it works to allow individuals who know the latter to use the former.

From Figure III-1 we can say that *an individual's ability to use an information system working on a specific application domain is largely determined by the validity of the model that this system has about the mentioned application domain*. We will call this assertion Principle of Usability.

The object-oriented paradigm, a "storm in a teacup" for some but a big turn in software conception for others, may help to build better models, so we can increase the usability of the information systems we make. Nevertheless, it lacks some concepts that are necessary in order to apply it to humanities disciplines, but we believe that it can be easily extended in the appropriate directions.

III.2. Methodology

The Fusion method, an industry-accepted methodology to build object-oriented information systems, has been adopted by us and modified to better fit our needs. We have proposed and published the Metis methodology, previously called the Adpated Fusion Method (MFA), that captures the best from Fusion and adds specific notation and process in the following areas:

- User-interface design and construction.
- Persistent storage design.
- Collaborative teamwork.
- User-documentation composition.
- System measuring.

IV. Information Models

Once we have developed a suitable methodology to build information systems, we can formalize the information often found during archaeological practice and document the models generated during the process called “Modeling” in Figure II-1.

IV.1. Abstract Classes

At a very high level of abstraction, we can consider three different types of objects:

- Locatable objects, the target of the archaeological practice.
- Documental objects, that document and describe locatable objects.
- Organizative objects, that give an organizative structure to support other objects.

Locatable objects can be furthermore classified as:

- Geographical objects, or geographical locations with no implicit archaeological interest.
- Archaeological record objects.
- Action objects, that represent actions on other locatable objects.
- Contextual analysis objects, with indirect archaeological interest because they offer information on other locatable objects.

Documental objects can be also subclassified as:

- Descriptive objects, that describe other objects in a more or less objective and timeless way, so they are *not versionable*.
- Valorative objects, that describe other objects in a subjective and time-related way, so they are *versionable*.

IV.2. Geographical Objects

They include points, lines and areas as geometric simplifications with no one-to-one correspondence between these classes and their potential substrates: for example, a river may be represented as a line or as an area, depending on scale and interest.

IV.3. Archaeological Record Objects

They include spatial, stratigraphic and material objects. Spatial objects often correspond to sites or site sets. Material objects include both structures and finds.

IV.4. Action Objects

They include aggressive and non-aggressive actions on other objects. Aggressive actions can be excavations or restorations, which usually involve a significant alteration of the object on which they act. On the other hand, non-aggressive actions include surveying and probing, which seldom modify their targets.

IV.5. Contextual Analysis Objects

Archaeological samples and descriptions of environmental conditions are subtypes of this kind of objects.

IV.6. Descriptive Objects

Many subtypes of descriptive objects can be considered: cartographic objects, terrain models, drawings, photographs, video clips, stratigraphic matrix diagrams, cards, etc. Furthermore, cartographic objects may be classified as maps, map layers and map entities.

IV.7. Valorative Objects

They include reports and state modifiers. Reports inform on other objects, while state modifiers change the valorative chain state (see *Theoretical Foundation*, p. 10) of their target.

IV.8. Organizative Objects

Resources (both human and material), functions, projects and tasks can be included as organizative objects.

V. Technological Proposal

This chapter develops the ideas underlying the box "Methodologies & Technologies" inside "Computing Domain" in Figure II-1.

V.1. The WSAP Architecture

The WSAP model defines a high-level architecture for information systems, specially useful in humanities applications. Its specificity lives in the introduction of the concept of representation at the architectural level: each letter in WSAP stands for a conceptual layer at which the system may be described:

- World. Captures and describes the entities and relationships of interest for the application domain.
- Structure. Includes the computerized referents that give access to the entities included in the World layer. A system may keep several referents for the same world entity, depending on time, purpose, etc. It is important to avoid functional semantics at this layer.
- Application. Involves the objects and relationships regarding the specific application intended for the information system; this is a functional description made on top of the structural one given by the Structure layer.
- Presentation. Includes all the objects necessary to display information to the final user and let her interact with the system.

Figure V-1 shows the general layout of an information system following the WSAP architecture.

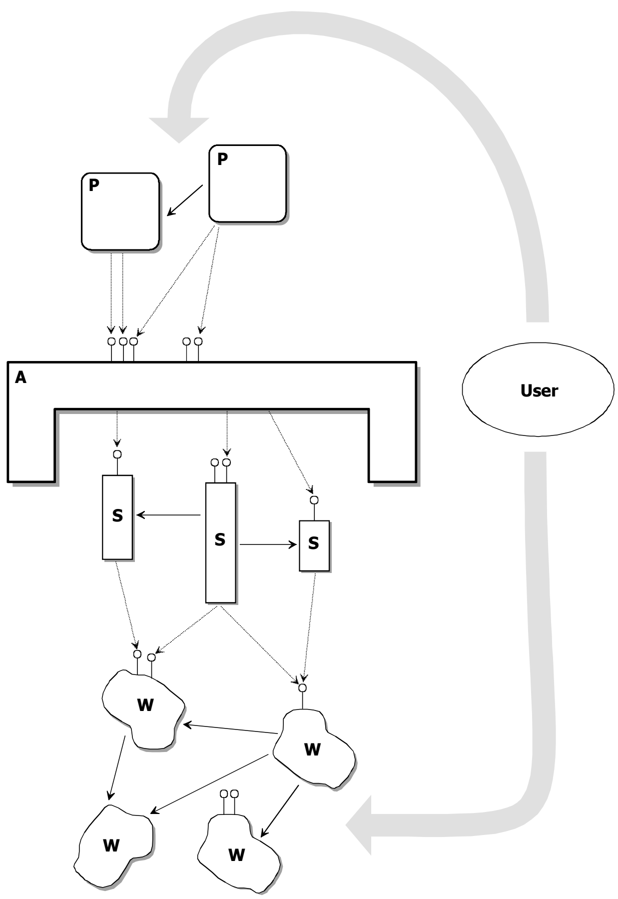

Figure V-1. Layers and relations in the WSAP architecture. The user knows the application domain ("knows" the world) and uses the information system through its presentation layer. See *Theoretical Foundation*, p. 13 for a explanation of these interactions.

The WSAP model implies the existence of a *representation manager* into the information system. When the user requests access to an application-domain entity (such as a river on a map, for instance) through the presentation layer, the representation manager must get a suitable referent from the application and structure layers. The same underlying entity might be represented in different ways depending on representation context (subjectivity, time), purpose or other factors. For example, the same river could be shown as a line on a map, as an area on a map, as a text description on a report, as an entry in a table, etc.

V.2. Object-Oriented Databases

The most currently used paradigm in database management systems is the relational one, which presents some lacks in expresiveness and is difficult to integrate with object-oriented programming languages. We think that object-oriented databases would make information systems much more affordable and usable.

A very important issue is that of metainformation. Metainformation can be defined as the set of referents in a system whose substrates are *also* inside the system. Although conventional relational database management systems provide metainformation in varying degrees, object-oriented ones would be able to offer a wider and richer metainformation scheme. Usually, information systems present metainformation mirrored at threee different levels:

- The database scheme "knows" its structure, be it in the form of tables and relations or classes and relationships.
- The processing logic of the system must also "know" the structure of the data it is using.
- Finally, the user interface must fit the same structure, presenting data organized in the same fashion.

This redundancy makes any change difficult and costly: for instance, adding a new attribute to a class in a system implies, with the conventional architecture, changing the database scheme, changing accordingly the processing logic, and updating the user interface so it shows the new attribute. A good use of metainformation can ease this work, if we build systems that only keep metainformation at one level. We have proposed that metainformation should live only at the processing logic level. The database should be *substructured*, or based upon very simple concepts such as slots (single data items in the form of a pair name-value, such as "Age: 32") and frames (sets of related slots, such as ["Name: César"; "Age: 32"; "Lives in: Santiago"]). The processing logic in the information system can access substructured data through metric layers that inject specific semantics on top of the data, tipically based in the concepts of *class* and *relation*; of course, different metric layers may be used by the same or different systems in order to get different views of the same data. Also, the user interface should avoid storing metainformation. *Emergent user interfaces* must be used, so nearly every presentational element (dialog boxes, toolbars, menu structures, etc) are dynamically synthesized at run-time from the system's unique copy of metainformation.

Finally, we believe that a suitable object versioning subsystems should be integrated into the database, seamlessly offering time-management services to upper layers.

V.3. Landscape Models

Geographical information systems (GIS) have been extensively used in archaeology with different degrees of success. We believe that a wrong approach is often used, because cartographic visualization and geographic analysis are functions that must be integrated into a more generic information system and not pursued as goals themselves.

Any information system built around the WSAP architecture can obtain cartographic referents in a similar way as it obtains textual ones (see *The WSAP Architecture*, p. 17), so we believe that an "implicit GIS" approach would be much more useful and coherent than the so often found "GIS-centric" alternative. The representation manager in the system would provide features such as, for example, automatically adding or removing detail from the cartographic referents depending on scale, purpose and user preferences.

V.4. Document Management and Cooperative Work

A lot of information is stored in documents, and good lexical, syntactical and semantic conventions must be established in order to guarantee intersubjective understanding. Lexical and semantic conventions must be achieved by constructing thesauri and integrating them with the information system, so, for instance, a search for the word "river" would also consider searching for "water stream". Syntactical conventions may be reached by using a structure-oriented document authoring system, which could be based on the

ISO standard SGML or on the much more popular and affordable XML. Document type definitions should be developed and consistently used, and presentation templates or stylesheets could be used for rendering documents to different devices (screens, printers, web pages, voice synthesizers for the hear-impaired, etc.). Document contents should be stored as part of the structure layer in the WSAP architecture, so they can act as substrates of representations within the information system and ber versioned and managed through time as any other objects.

Also, providing the appropriate documents to the appropriate people just in the moments they are needed could be the goal for a workflow management system built around the same premises.

VI. The Archaeological Information System

The SIA+ (Archaeological Information System, enhanced version) is an integrated system developed by the author for the Laboratory or Archaeology & Cultural Forms (LAFC) of the University of Santiago in Spain. It has been in use for several years now, storing information on hundreds of thousands of archaeological entities and supporting the daily work of over twenty people.

VI.1. Functional Description

The first version of the SIA (Archaeological Information System) was conceived in spring 1993, and involved storing and managing information on entities such as sites, excavations, structures, finds, drawings, valorations and projects. It should be able to serve several users simultaneously and work on affordable computers. In autumn 1993 the system started to work, and a year later in contained data on 500 sites, 30.000 finds, 1.100 structures among other things. It also had search and reporting tools as well as cartographic export capabilities.

Two years later, the SIA could not keep up with the growing needs at the LAFC, so a total re-write was considered. The SIA+ was born on 1997, and included over 32 different types of entities, a new user interface, work contextualization capabilities, navigational searches, data entry tools, a new highly customizable report generation tool, and integrated cartographic visualization. Currently, the SIA+ manages information on over 3.400 sites, 120.000 finds, 2.400 structures, 11.800 photographs, 6.700 valorations and 170 projects.

VI.2. Theoretical and Methodological Analysis

The application domain was modeled using object-oriented methodologies and techniques, so class models and event models were obtained. This conceptual framework has been cross-checked a number of times within different contexts, and still stays valid. Cartographic information was included through the use of the CVS technology (see below), so geographic substrates were linked to cartographic referents into the system through a representation manager. Versionable entities were considered, and a very simple time and subjectivity management subsystem was designed. Finally, the theoretical principle of contextual interaction was used to give the system work contextualization capabilities.

VI.3. Technological Analysis

The SIA+ stores data into a relational database. Mapping class diagrams to a relational scheme was a difficult task, but some strict conventions helped. Microsoft Visual Basic and Microsoft Visual C++ were used to develop different components of the system. The Microsoft Windows NT platform was chosen because of its wide acceptance, affordability and wide range of tools in the market.

VII. Consequences

Four areas must be further developed in order to make us able to build information systems in the propsed ways:

- Substructured databases and access metric layers.
- Object versioning.
- Cartographic visualization.
- Document management.

All four areas are strongly interrelated (see *Technological Proposal*, p. 17) but may be developed within independent efforts.

Having defended that human understanding happens thanks to shared intersubjectivity among communicating individuals, we must try to ellucidate whether apparently existing technologies are in fact real or just water-mirrors produced by a too long journey through a too barren desert. Let us not confuse products with technologies, or foundations with results. Perhaps this thesis does not offer results nor products, but it surely provides a good amount of foundations and technological proposals that, with enough time and resources, will be able to render products and results with higher quality and more solid than current ones. This is a long run, and we will only get to the finish line by being honest and patient.