A MODEL AND A SYSTEM FOR CULTURAL HERITAGE: THE CASE OF THE HAT INDUSTRY

Feliz Ribeiro Gouveia - *Fernando Pessoa University, Faculty of Science and Technology*
Multimedia Studies and Resource Center (CEREM)
Full Professor | E-mail: fribeiro@ufp.pt

Sérgio Lira - *Fernando Pessoa University, Faculty of Social and Human Sciences*
Assistant Professor | E-mail: slira@ufp.pt

Abstract

This paper describes our current work in the area of Museum Information Systems, namely for crafts and industry museums, where the main concern is to preserve and document not only material artefacts but also oral traditions, know-how, processes, cultural and social contexts, local history, sounds, tactile experiences and smells. Our case study is a Hat Industry Museum, which is being created in the North of Portugal, at the heart of what was, in the last century, the biggest hat-manufacturing center. As such, the Museum will be also a tribute to the social and economic life of a whole region. We propose and describe a conceptual approach and an information system for describing non-material heritage and its application to our case study.

Resumo

Este trabalho descreve um projecto na área de Sistemas de Informação para Museus, especificamente para Museus de Indústria, onde existe a preocupação de preservar e documentar não só património material, mas também tradições orais, experiências, processos, aspectos sociais e culturais, história, sons, e experiências sensoriais. O nosso caso de estudo é o Museu da Indústria da Chapelaria em S. João da Madeira, em criação no coração da região que foi o maior centro industrial da chapelaria em Portugal. O Museu pretende ser um testemunho da vida económica e social de toda uma região. É feita uma proposta de um modelo conceptual e de um sistema para representar património imaterial, e é descrita a sua aplicação ao caso de estudo.

1. Introduction

This paper describes our current work in the area of Museum Information Systems, namely for crafts and industry museums, where the main concern is to preserve and document not only material artefacts but also oral traditions, know-how, processes, cultural and social contexts, local history, sounds, tactile experiences and smells. Our case study is a Hat Industry Museum, which is being created in the North of Portugal, at the heart of what was, in the last century, the biggest hat-manufacturing center. As such, the Museum must preserve not only the material heritage, be it machines, hats, documentation, tissues, and other artefacts, but also non material heritage such as the effects on the lives of the people and the whole region. Given that most of the industries and the region suffered a profound economic impact, and most of the history and the people who made it are disappearing, there is an urgent need to preserve their heritage. In this context, oral traditions and history, mostly from whole families of workers, are among the most important intangible pieces of this industry museum.

From the museum perspective, the system is very simple to use. We have an interface that enables authorised users to input and/or manipulate data and another interface that enables the retrieval of information. The first one "builds" the individual record for each item as it is being completed, using the inheritance scheme of the model. The second one offers complete search possibilities on what concerns the fields to be searched, or the contents of those fields. And as the system works using an Internet browser, the search can be made locally or on a remote Internet access. In the future, as presently under development project aims at, the museum's web site will facilitate the access to the inventory system wherever researchers are.

The rest of the paper is organized as follows: in the next section we describe the case study and the methodology being used to identify, collect, preserve and organize both material and non material heritage. In section 3 we review existing conceptual modelling approaches as being applied by the Museum and other communities. We emphasize also current trends in information representation, communication and access techniques and technologies that are relevant. In section 4 we briefly describe the architecture of the system. Section 5 describes the conceptual model we use, and sections 6 and 7 describe interaction issues. We close by discussing what we have learned so far, and what we hope to contribute to documenting non-material heritage.

2. The hat industry museum

The Hat Industry Museum is being created in the region that was once the capital of Portugal's hat making industry. As such, the hat industry had a deep impact upon the lives of generations of thousands of people. Culture, habits, social life, crafts, was heavily influenced by the hat industry. Most of the factories are now closed, and pieces witnessing their prosperity are lying abandoned. The challenge is to scientifically and systematically gather, register, classify this material heritage, and, at the same time, register what we can consider non-material heritage, such as traditions, habits, know-how, skills and histories. Using an old factory as its main building, the Museum will recreate the sounds, the feeling, the processes, and will keep alive the heritage of the region. The Museum will use the Internet to make available all this information, in a standard as possible format, which prompted us to start developing a conceptual model for non-material heritage. An initial introduction to the Hat Industry Museum project is given in (Lira, Menezes, 2001; Gouveia et al., 2002).

At the Hat Industry Museum we have very different kinds of documents. The main groups are the documents that once belonged to the factories (mainly administrative and commercial documents) and the interviews with former workers that are being gathered as the result of anthropological research. But we also have very important collections of photographs, advertising material and medical records, among other categories. Most of the physical objects that we have collected from existing or already closed Hat Manufacturers belong to the following categories:

- heavy equipment;
- parts and supplies of machinery;
- raw materials;
- finished products (hats, hats accessories);
- administrative documentation, advertising materials, price lists;
- process documentation, machine operating manuals;

A great deal of effort was spent however registering oral statements of former actors, be it workers, managers, clients, people remembering histories; registering uses of equipment and materials, which are not used anymore, is also important, as most of the skills were non documented, and people learned at the work floor how to make hats. In order to understand the material evidences of the hat industry, as it was a century ago, it is of vital importance to record and make accessible all the information still available.

Oral statements are being recorded using high quality media, such as minidisks and miniDV tapes. The quality and long-life perspective of the records is crucial as these records will be of fundamental importance in the archive and documentation center of the museum. They will be the basis of permanent exhibition interpretation as well as essential material to internal research projects. They will also be available to be used by external researchers, locally or on-line. As the museum seeks a high level of integration with the regional community, these records will be used in projects developed in partnership with local schools.

This documentation module is able to accept text, sounds, images and videos. On what concerns script documents, the least demanding records are the ones that only register the contents of the document. In another level are those records that register the exact text of the document, transformed in word-processed text using an Optical Character Recognition program. At the higher level are those records that register a facsimile (scanned) version of the document.

The criteria for classifying documents in these three categories are pre-established and combines three main aspects: relevance of the document in terms of understanding the history of the hat industry, existence of relevant graphic issues, and aesthetics. Oral testimonies and videos are also treated as documents, and records will have the proper link to the sound or video clip. Images will be shown in low resolution and a link to high-resolution sample will be available. Whenever different documents have some kind of connection, records will make that reference, enabling a complete research.

Another demand was that documents should be connected to other inventory issues, as material artefacts, machines and buildings. One of the main goals of the system is to enable the retrieval of complete information on any hat industry subject, whatever the beginning of the search.

3. Conceptual information modelling

As Museums, and other entities, move from dealing with tangible, material entities such as paintings, works of art, crafts, machines, to more intangible entities such contexts, implications between entities, history, traditions, there is a need for more semantically powerful and meaningful information models. There is also a shifting role that Museums, Libraries, Archives and similar institutions are playing now, with the possibilities the Internet and the Web have opened up.

Although conceptual modelling techniques have been proposed in the late seventies, their application by the Museum and Information Sciences community has had to overcome the analysis of the huge work and investment already made in cataloguing standards, electronic records standards, and Information Systems. Recently, the Documentation Standards Group of the ICOM/CIDOC proposed an Object-Oriented Conceptual Reference Model (ICOM/CIDOC, 1999), based on the notion of Ontologies, as proposed by Gruber (1993). The Conceptual Reference Model (CRM) is itself an evolution of the CIDOC's Relational Model (ICOM/CIDOC, 1995; Reed, 1995).

The CRM intends to be an ontology for Museum data, and as such several mappings and translations were developed. For example, a mapping between the Consortium for the Computer Interchange of Museum Information data model and the CIDOC Reference Model is given in (Kody, Sledge, 1995). Other mappings are proposed by Doerr (2002). As it is being considered mature, the CIDOC CRM (version 3.4.9) is now being proposed as an ISO standard (CIDOC, 2004).

The International Federation of Library Associations and Institutions proposed also a conceptual model to deal with bibliographic information, recognizing the pressing need of providing a clear and shared understanding of "what it is that the bibliographic record aims to provide information about, and what it is that we expect the record to achieve in terms of answering user needs" (IFLA, 1998). The IFLA FRBR (Functional Requirements for Bibliographic Records) model breaks with the file-oriented family of MARC models in that is more semantically oriented and allows more meaning to be added.

Other initiatives worth mentioning are the Open Archives Initiative (OAI), which promotes a protocol for interoperability between different systems based on metadata extraction (OAI, 2003).

We based our work on semantic constructs, such as ontologies, that define the concepts and terms in a given domain in a way that can be read by non-specialists. An ontological approach has several benefits:

- it is a high-level description of a domain, independent of implementation details; such a conceptualization can say that a book is authored by an author, leaving to the implementation the details of deciding how to name the entry or what cataloguing standard use; Ontologies can be translated to different implementation systems and formats;

- several institutions can share and agree upon an ontology (see previous example), regardless of the specifics of their implementations;
- besides being shared, ontologies can be reused; once a Time or Units of Measure ontology has been created, several institutions can reuse it without them having to define themselves the ontology. The Ontolingua project at Stanford University is an example of a shared repository of Ontologies, written by different contributors (Farquhar et al., 1996).

Based on our previous experience in Document Management and Information Systems, concepts such as inheritance, specialization, generalization, and encapsulation can greatly enhance the process of documenting and cataloguing, and accommodate easily modifications, evolutions, and mappings to other standards or formats.

4. Architecture of the system

The system is based on a client-server architecture using TCP/IP, the Internet protocol, and a relational database management system, as shown in the next figure:

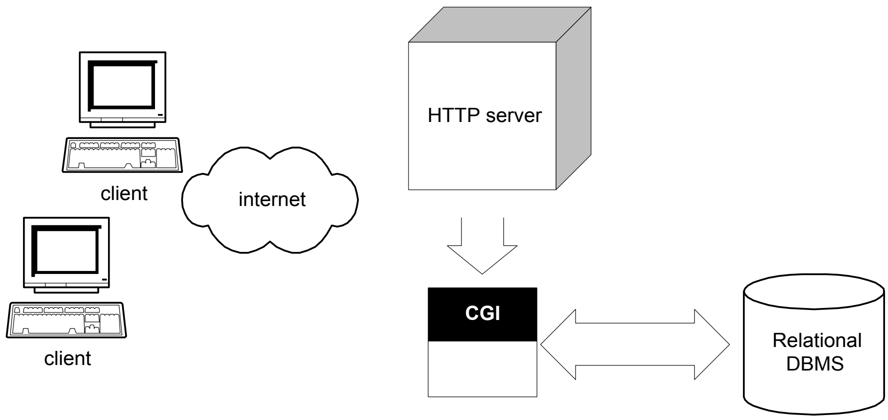

Figure 1: Architecture of the system.

Clients connect to the system using a standard browser via a network. Browsers must be JavaScript enabled and support at least HTML 4. For some applications, a secure connection is established using the Secure Sockets Layer protocol. The interface presented to the users introduces a minimum overhead, by reducing image and graphic elements, and by not using frames as this can bring some problems with some browsers.

The software architecture of the system uses a relational-object layer to present to the user an object-oriented view of the data, and to store it in a table format in a relational database. This is a common approach when we want the expressive power and the flexibility of the object model, and the convenience (even in cost terms) of working with a relational database. The CIDOC CRM also follows the object-oriented approach. A report from the Stanford University (Paepcke et al., 1999) lists some of the limitations of flat (e.g. Unimarc) and relational models to implement digital library services of the kind we discuss in this paper.

Classes with instances are mapped to tables, which contain all attributes, including inherited ones. Relationships are mapped to tables with just two object identifiers. Although this is not the most efficient strategy (e.g. see Agarwal et al., 1995), it allows for most of database accesses to be made with only one join operation. Future improvements could distinguish between different types if relationships and decide different strategies for their storage in the relational database. If performance issues arise, an object-oriented database could be used in the future (Kim, 1990).

We defined a set of rules to present information to the user by taking advantage of the object-oriented model: browsing, navigating and searching information follows the structure of the model. Similar approaches have been proposed by Paton et al. (1994).

5. The object-oriented model

We follow closely the OMG model (OMG, 1992). The object-oriented model we use has the following characteristics:

- Classes, with or without instances;
- Classes have a Name, Attributes, and Relationships to other classes;
- Attributes have a Name, a Type, and can be optional;
- Relationships have a Name, a Multiplicity, and a Type which is a set of classes;
- Single inheritance. There is a 'root' class which is the top of the class hierarchy;
- Instances belong to one single class.

Attributes and relationships can be redefined in the sub-classes. This model avoids multiple inheritance as it can be difficult to implement in practice, and to work with. Moreover, the redefinition of attributes and relationships would turn the model even more complex if multiple inheritance was allowed. This contrasts to the CIDOC CRM, which uses multiple inheritance.

Relationships such as 'part-whole' (see for example, Barbier, 2003) were not considered in this version of the system. Other types of associations will certainly emerge in the future (see for example Blaha and Premerlani, 1998, Liu and Halper, 1999, Motshnig-Pitrik and Kaasbøll, 1999).

The following figure shows a simplified Entity-Relationship model of the object-oriented model we use. For simplicity we omitted the relationship's multiplicity information; the meaning and the reading of the relationships should be clear from the model.

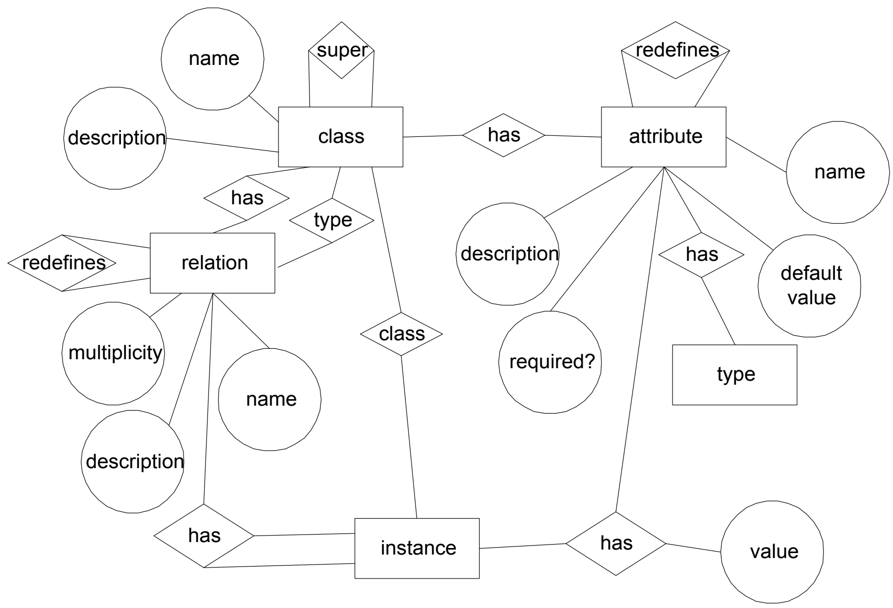

Figure 2: The E-R model of the system.

Classes, attributes, relationships and instances have privileges for Editing, Reading, Creating and Searching. Classes and instances have an owner: the instance owner is the creator of the instance.

Privileges are specified for the owner, the group of the owner, and for the world, much like Unix privileges. We have found this a very convenient way of formatting information to be viewed by different users; for example, some documents contain sensitive, medical, data about employees which is only shown to members of some groups, but are hidden from other users and from the general public. This means that some attributes and relationships are only readable by some people. Other approaches (Adam et al., 2002) use more complex privilege models, based on the content of the objects, the user, and are specified for a single object or for sets of objects. The DSpace (DSpace, 2004) digital library systems uses an extended version of our privilege model, including delete privilege and collection-level privileges.

Instance privileges were also found to be useful. In some cases, other users should not view instances created by some users; this is the case of personal folders for example. In other cases, we want an instance to be edited by the members of some group, but not by other users. For example, users belonging to the ‘management’ and ‘cataloguing’ group could have different privileges over the same classes and instances. Users belonging to the ‘administrative’ group should not be able to edit cataloguing information, and users belonging to the ‘cataloguing’ group should not be able to edit administrative information.

Typical privileges are given as 'F75', meaning full privilege (edit-read-create-search) for the owner, read-create-search privilege for the group and read-search privilege for the world. Privileges are checked first at class level, then at instance level, and then at attribute and relationship level.

The following figure shows a small part of the class hierarchy for the Hat Industry Museum. The system currently has around 140 classes, over 400 attributes, and 50 relationships.

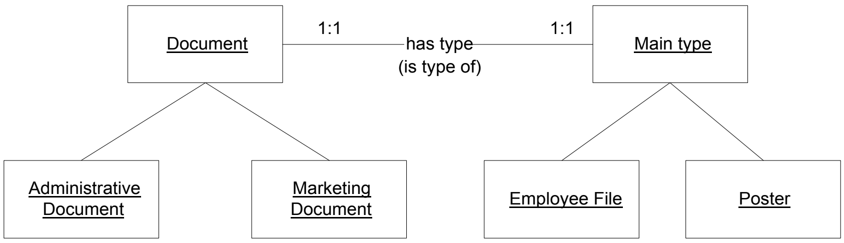

Figure 3: Part of the class hierarchy

In figure 3, relationship 'has-type' links a 'document' instance with a 'main type' instance and it is the most used relationship: every instance of 'document' has a relationship with one instance of 'main type'.

6. Navigating the model

The following figure shows the graphical appearance of such an instance. Gray bars are used to show relationships. When instances in the relationships are shown, the name of the class is also shown just after the bar.

Figure 4: Viewing an instance of a document.

In the left side of the figure we show the conceptual relationship between the instances. In the right side we show what is presented to the user.

In this case, the user can see the instance of document, and also the instance in the relationship 'has type' ('acto principal'). This is because this relationship is declared as 'inline', which means that the instances are also shown in the screen. In the figure below we show an instance that has relationships declared as 'link', which means that only a link to the instances is shown:

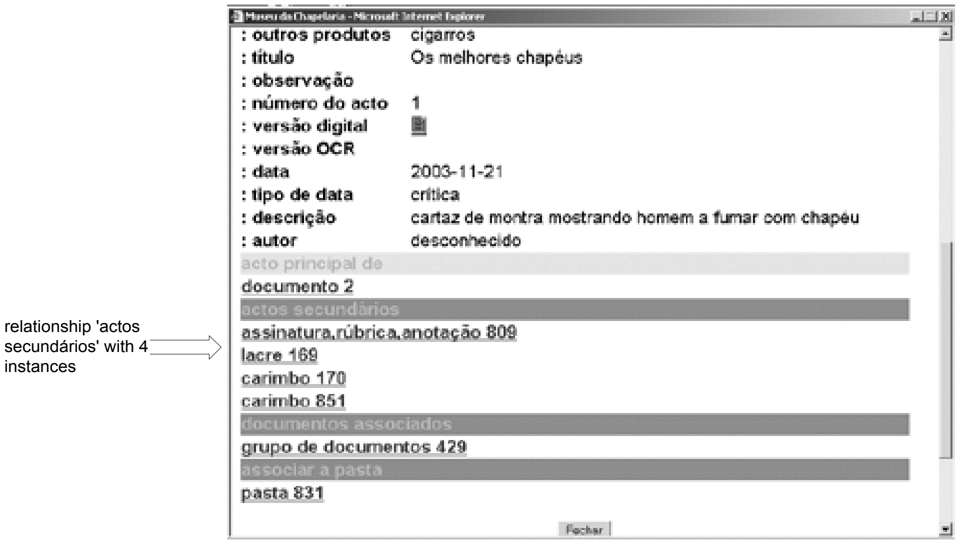

Figure 5: Instance with relationships shown as links.

When creating instances, relationships can also be defined as accepting only new instances, existing instances, or both cases.

For example, when a user is creating a new instance of 'document' he must create a new instance of 'main type': it would not make sense to use an existing 'main type' instance (for example an existing 'medical record'). In some cases, it makes sense to use existing instances, as it is the case when creating an association between a 'document' and a 'folder': we can create a new folder, or use an existing one. The figure below shows an interface for creating an instance of document where both cases are shown:

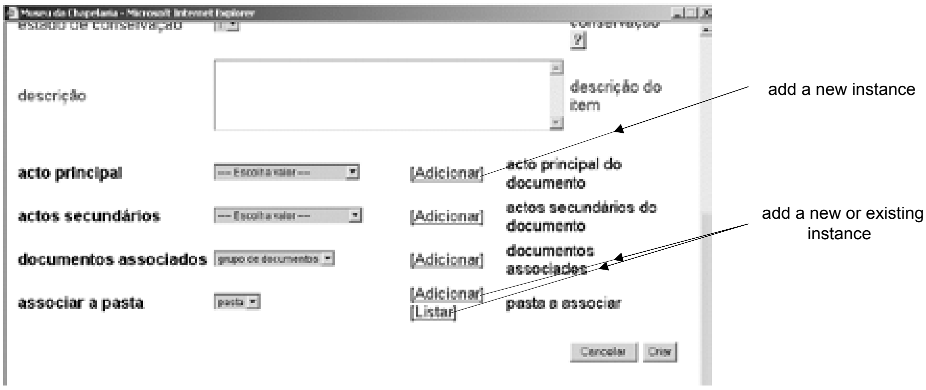

Figure 6: Adding instances to a relationship.

How to add instances to a relationship is a constraint that is defined at class definition time. Besides this and the multiplicity constraint, no other constraints are used, although we are planning to add a set of checking constraints, much like integrity constraints in a relational database.

7. The search interface

The search interface presents the user with a view of the model, from which the user can select attributes to list in the results, and restrictions to enter, as shown in the figure below.

Figure 7: The search interface.

Class attributes and relationships are shown according to the privileges defined for the current user. If an attribute can be seen and searched, a text box is shown next to the attribute. If the user checks the box near the attribute name it means he wants to see that attribute's value in the result list. Search is performed as follows:

- Numeric values and dates can be searched using numeric operators (see for example the operator's option list after the 'data de inventário' attribute);
- String and text attributes use a partial match; if more than one word is given, a logical 'or' is performed.

In a future version, the user will have the option to perform exact text matches, and to use logical 'and' of words. Another interesting feature, already implemented, is to check for all attributes matching their values or to have at least one attribute matching its value.

SQL code is dynamically created from the interface, generating temporary tables if necessary. Dealing with issues such as joining relationships, as described in (Lee and Wiederwold, 1994) was restricted to both sides of the relationship matching the instances; in the future, outer joins would allow for more powerful searches, specifying for example that an instance must or must not belong to a given relationship.

Relationships are also shown subject to the same reading and searching privilege restrictions. If a user is allowed to see and search in the relationships these are shown as in figure 7. The user can expand any of the relationships, thus seeing the attributes and other relationships belonging to that class. Figure 8 shows the expansion of relationship ‘main type’ (‘acto principal’).

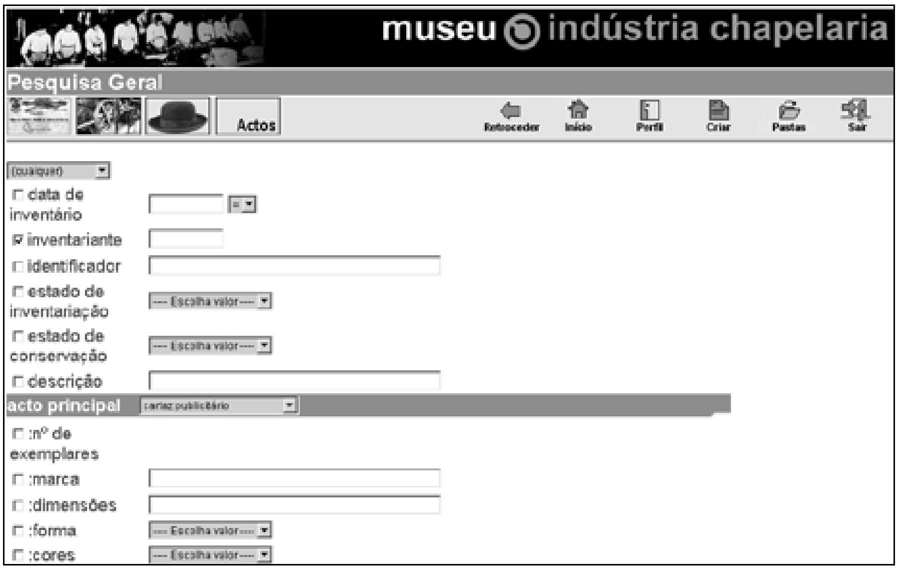

Figure 8: The search interface.

Searching is thus equivalent to navigate the object-oriented model of the system. Following (i.e. navigating) relationships is achieved by collapsing and expanding the links shown in the interface.

Figure 9 shows the same search window as seen by a user with fewer reading and searching privileges: only two attributes are shown and the relationship ‘actos secundários’ is not shown. Note that both attributes are checked, so that the result lists two columns, one for each set of values.

Figure 9: Listing results.

Any instance in the result list can be seen by clicking in the link, if the user wants to browse for more information. Note also that in the example, the user is searching for instances of 'any' document having the word 'cartaz' in the 'descrição' (description) attribute.

8. Concluding remarks

When we began this project, we aimed at establishing a system that should respond to the CIDOC's CRM standards (e.g. Reed, 1995, ICOM/CIDOC 1999). But we also wanted a system that would be easy to use and adaptable. This system can be updated whenever the Museums needs: different fields and new types of documents can be created very easily. As a consequence, we do not need to have all the documents' categories and types, and all the descriptive fields of each category or type, pre-established. As the system is being used, we have been experiencing the advantages of our approach, and gathering knowledge for improvements to future versions of the system.

As our documentation will develop, our system will be updated. Finally, the system can be used in any language, as the names of the fields are all gathered in internal tables of the system; as a consequence, they are very easy to translate; it is not necessary to change the names of the fields in each individual record, as it inherits its characteristics from the system's scheme.

While developing the conceptual model of the Hat Industry Museum, priority was given to integration with existing efforts and standards so that future uses could be accommodated. Ontologies can be shared, are implementation independent, and translators can be written to different formats; interrogation languages can also be designed and used — see e.g. the approach in (Kody, Sledge, 1995) for the Z39.50 protocol — as long as the translation is done locally. We found that we needed a rich information model so that the information could be registered, and accessed. For example, oral information refers to how to cut or prepare a raw material, what is this tool for, or what skills are needed for a given process. We hope to keep watching and contribute to future developments in Museum Information Systems, as conceptual standards and recommendations are the key to inter-operability of existing and future systems.

Bibliography

Adam, N., Atluri, V., Bertino, E., Ferrai, E. (2002). A content-based authorization model for Digital Libraries. *In: IEEE Trans. on Knowledge and Data Engineering*, vol 14, nº 2, March/April, pp. 296-315.

Agarwal, S., Keene, C., Keller, A. M. (1995). *Architecting Object Applications for High Performance with Relational Databases*, High Performance Object/Relational Applications, August.

Barbier, F., Henderson-Sellers, B., LeParc-Lacayrelle, A., Bruel, Jean-Michel (2003). Formalization of the Whole-Part Relationship in the Unified Modeling Language. *In: IEEE Trans. on Software Engineering*, vol 29, nº 5, May, pp. 459-470.

Blaha, M., Premerlani, W. (1998). *Object-Oriented Modelling for Database Applications*, Prentice-Hall.

CIDOC (2004). "CIDOC Conceptual Reference Model. Official Release of the CIDOC CRM", [online] available at <http://cidoc.ics.forth.gr/official_release_cidoc.html>.

CIMI (1999). *A model for Museum Information Management.* CIMI Consortium Integrated Information Management Working Group.

Doerr, M. (2002). Mapping a Data Structure to the CIDOC Conceptual Reference Model. *In: Proc. of Computer Applications and Quantitative Methods in Archaeology*, CAA2002, Heraklion, Crete, Greece, April 2-6.

Dspace (2004). "DSpace System Documentation: Functional overview", [online] <http://www.dspace.org/ technology/system-docs/functional.html#data_model>, HP and M.I.T.

Farquhar, R. Fikes, J. Rice (1996). *The Ontolingua Server: a tool for collaborative ontology construction*. Technical Report KSL-96-26, Knowledge Systems Laboratory, Stanford University.

Gouveia, F., Lira, S., Menezes, S. (2002). A conceptual approach for non-material heritage: the case of the Hat Industry. *In: Proc. of Preserving cultures: documenting non-material heritage,* CIDOC2002, International Council of Museums, Porto Alegre, Brazil, 18-20 September.

Gruber, T. (1993). *Towards principles for the design of Ontologies used for knowledge sharing* Technical Report, Knowledge Systems Laboratory, Stanford University.

ICOM/CIDOC (1995). *Graphic Data Model*. ICOM/CIDOC Data Model Working Group.

ICOM/CIDOC (1999). *Definition of the CIDOC object-oriented Conceptual Reference Model.* ICOM/CIDOC Documentation Standards Group.

IFLA (1998). *Functional Requirements for Bibliographic Records*, Final Report, UBCIM Publications – New Series vol. 19, IFLA Study Group on the Functional Requirements for Bibliographic Records.

Kim, W. (1990). *Introduction to Object-Oriented Databases*, MIT Press, Cambridge, Mass.

Kody, J., Sledge, J. (1995). A User Model for CIMI Z39.50 Application. [online], available at <http://www.cimi.org/public_docs/Z3950_app_profile_0995.html>. Visited the 13th March 2002.

Lee, B. S., Wiederhold, G. (1994). Outer Joins and filters for instantiating objects from relational databases through views. *In: IEEE Trans. on Knowledge and Data Engineering*, vol 6, nº 1, February, pp. 108-119.

Lira, S., Menezes, S. (2001). The Hat Industry Museum of S. João da Madeira (Portugal). Museological Review, Leicester, Department of Museum Studies, vol. 7, p. 68-71, available at <http://www.ufp.pt/~slira/Projectos/museudechapelaria/museologicalreviewartigocompleto.htm>.

Liu, L., Halper, M. (1999). Incorporating Semantic Relationships into an Object-Oriented Database System. *In: Proc. of the 32nd Hawaii International Conference on System Sciences*.

Motshnig-Pitrik, R., Kaasbøll, J. (1999). Part-Whole relationship categories and their application in Object-Oriented analysis. *In: IEEE Trans. on Knowledge and Data Engineering*, vol 11, nº 5, September, pp. 779-797.

OAI, The Open Archives Initiative Protocol for Metadata Harvesting. Open Archives Initiative, [online] <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>. Online version of 2003/21/02.

OMG (1992) *Object Management Architecture Guide*, 2nd ed., R. Soley ed., Object Management Group Inc.

Paepcke, A., Baldonado, M., Chang C.-C., Cousin, S., Garcia-Molina, H. (1999). Using distributed objects to build the Stanford Digital Library Infobus. *In: IEEE Computer*, pp. 80-87, February.

Paton, N. W., Al-Qaimari, G., Doan, K. (1994). On interface objects in Object-Oriented Databases. *In: 12th British National Conference on Databases*, D.S. Bowers ed., Springer-Verlag, LNCS 826, pp. 153-169.

Reed, P. A. (1995). *CIDOC Relational Data Model. A Guide*, ICOM/CIDOC.