

*To appear in the proceedings of MDA Conference 1991  
European museum documentation strategies and standards.  
3-6 September 1991 - Canterbury, England*

## **Cataloguing art objects: a comparison between French and Italian standards**

*Oreste Signore*

CNUCE, Institute of CNR - via S. Maria, 36 - 56126 Pisa (Italy)  
Tel. +39 (50) 593201 FAX +39 (50) 576751  
E.mail: EARN.bitnet ORESTE@ICNUCEVM.CNUCE.CNR.IT

### **A - Abstract**

The methodologies adopted by the national cataloguing institutes of France and Italy are briefly compared, pointing out similarities and differences, from "intellectual" and "technical" point of view. A greater emphasis is put on the structuration of data as a mean for the exchange of the data.

### **A - Introduction**

In the aim of starting a close cooperation between italian and french governments, having as goal the possibility of exchanging the data pertinent to the italian works of art presently located in France, and vice-versa, a comparison has been made of the standards adopted by the two institutions, examining the possibility of defining an exchange format. Even if more attention has been payed to the data regarding paintings, the two different approaches to the definition of a classification schema have been compared from a general point of view. In the following, a brief description of the basic assumptions made in the two different countries is given. The two approaches are

compared from the point of view of the data model, the controlled vocabularies, and the iconographic description systems adopted.

Finally, the importance of the agreement at the conceptual level as a basic requisite for an effective exchange of data is discussed.

## **A - The approach of the Italian Catalog**

The first organisation of the Italian Catalog was based on a manual approach. Therefore, each object was described by a typewritten card, following some general rules which greatly influenced the definition of the fundamental principles of the cataloguing schema.

The most relevant issues were:

- the identification of a reduced set of different cards, corresponding to different types of objects (art objects, archaeological objects, drawings, architecture, gardens, historical centre, etc.);
- the grouping of the information in several very general categories, like author, location, material, historical info, etc.;
- the topological arrangement of the catalogue cards.

On the other hand, it has to be pointed out that the cards were conceived for human usage, and not for automated treatment. As a matter of fact, around the mid '70, attempts were made to directly transpose the cards into information retrieval systems, but the results were very disappointing: as far as the quantity of data was increasing, the precision and recall factors were going down.

It soon became evident that the unsatisfactory results could not be ascribed to the peculiar systems, but that a rethinking of the entire cataloguing schema was necessary, in light of the constraints imposed by the automated treatment of the information.

The definition of the data structure started from the identification of the central role played by the object. It was agreed to define a classification schema based on three different kinds of objects: simple objects, complex objects, aggregation of objects.

A simple object is an object such that all his attributes are pertinent to the whole object, and no components which may themselves be considered cataloguing objects may be identified.

A complex object may be either a simple object whose parts, physically or conceptually separable, exhibit some interesting peculiarities as cataloguing objects, either a set of objects which may be referred by a specific name.

The aggregation of objects arise when several objects are correlated on the basis of some conceptual criterion, but no name exists which identifies the aggregate.

It is obvious that the components of a complex object may be either simple either complex objects, and so is as far as the aggregate objects are concerned.

It is worthwhile to note that a specific object belongs to the different categories only on the basis of the quantity and the type of information: no list exists that specifies if a particular kind of object must be considered simple or complex or aggregate. The proposed model only establish a classification model, that is, the type of relationships that must be specified between the objects (a component of a complex object is an object itself), and the criteria of inheritance of the properties.

The adopted approach was essentially based on the standard methodologies currently in use in the conceptual database design area. As it is well known, the most popular approach is based on the Entity-Relationship model proposed by Chen (1983), and the conceptual model is independent on the software and the hardware environment.

The first step has been the identification of the "basic" entities, like object, author, location, and so on. The identification of the relationships between these entities has been taken as the second step. This process lead to a simple, consistent model, were the object was playing a central role.

In order to have a normalized vocabulary, some "authority files" were defined.

Even if the Entity-Relationship model has been recognized as a very effective communication mean between end users and designers, it was found more productive to use an alternative way of representing information, that is, to switch to the conventional "cataloguing card" format.

From this point of view, the following choises were made:

- the information has been subdivided into small, semantically well defined, chunks;
- these chunks may be either a field, either a subfield of a structured field;
- each field may be defined as simple or structured;
- each field may be defined as repeating or non-repeating;
- each subfield may be a repeating or non repeating subfield;
- fields, either structured or unstructured, may be grouped into "paragraphs" in order to allow multiple occurrences of a set of fields.

At first glance, the proposed model may appear just another "flat file" schema, with a large number of fields, but anyone familiar with the database design methodologies will

easily recognize that, generally speaking, entities have been mapped onto paragraphs, (multivalued) attributes onto (repeating) fields, aggregate attributes onto structured fields, as reported in Signore, Pardini and Trasacco (1986).

It is also evident that the identification of a sequence of fields, with the characteristics of being repeatable and/or groupable, and references to "authority files", may be seen as the "linearization" of a non linear text.

Last, but not least, an effort has been made in order to maintain consistency between different cultural areas, so that semantically equivalent fields are identified by the same tag.

The guidelines followed in the definition of the standards are reported in Papaldo, Ruggeri, Gagliardi, Matteucci, Romano and Signore (1986) and in Signore (1985).

A complete definition of these standards may be found in the publications of the ICCD: D' Amadio and Simeoni (1989), Massari, Prosperi Valenti Rodinò, Papaldo and Signore (1988), Papaldo, Ruggeri, Gagliardi, Matteucci, Romano and Signore (1988), Parise Badoni and Ruggeri (1988).

## **A - The approach of the Inventaire**

The approach taken by the Inventaire is substantially based on the information retrieval systems techniques. The decision to adopt this kind of techniques was taken since the beginning, i.e. in 1964. A first step has been the definition of methodological tools, like dictionaries and classification schemas. Successively, since 1977, the computers began to be massively used. Toward the end of '80, microcomputer usage was becoming an important issue.

All the informations are arranged into fields, and fields are grouped into general categories, like geographical data and localization, object, chronology, cultural definition, technical data, analytical data, administrative data.

Looking in more detail to the classification schema adopted in the database Joconde, it appears evident that each object is treated as a single. If an object is part of a set of objects, references are present to the other elements of the set .

The complete cataloguing schema has been conceived having in mind the facilities offered by the adopted software (MISTRAL). This means that an extensive use has been made of the peculiarities of the system, like the possibility of inserting special delimiters in the fields.

Each field has been defined either as having a controlled vocabulary (or thesaurus), or as a free text field. From the usage point of view, inverted indexes are built on

vocabulary controlled fields and on some free text fields. On the remaining free text fields only a "secondary" or sequential search is possible, and this is allowed only on the subset of the database selected by a previous query based on inverted fields.

## **A - The two approaches compared**

First of all, it is worthwhile to note that at first glance the two approaches do not differ very much. Both present the data in the conventional format of a "cataloguing card", and organize the information into small fields, grouped together.

From this point of view, the definition of an exchange format seems to be reduced to the identification of the correspondences between the different fields.

It should be noted, however, that there are some basic differences, either from the "intellectual" either from the "technical" point of view.

### ***B - The data structures***

The first basic difference is that the Italian approach is pursuing the objective of defining a cataloguing schema which is independent on the software and hardware environment, while the French schema has been greatly influenced by the underlying software. This difference is probably due to the fact that the French standards were initially defined when the database design methodologies were just at the beginning stage. The Italian schema is much more structured than the French one. This is a consequence of some previous experiences, where the adoption of delimiters at the data entry stage resulted in some inconsistencies, while it is very simple to automatically put together information subdivided into smaller elements.

In the Italian model, a great emphasis is put in the definition of the different types of relationships between objects, leading to the definition of the different types of objects: simple, complex, aggregate. It has to be noted that a component of a complex object is seen as an object itself, and inherits some properties from the "father object". The aggregation of objects leads to the compilation of a cataloguing card that accounts for the general properties of the aggregate. In conclusion, we may envisage two different types of relationships between objects: a "vertical" relationship (complex objects) and a "horizontal" relationship (aggregates). In the French model, the aggregate object concept is missing. Finally, it has to be noted that this classification model does not imply any hypothesis on the type of system that will be used. In fact, it has been

successfully mapped onto Information retrieval systems as well as onto Data base management Systems of various types.

### ***B - The vocabularies***

There is a complete agreement on the fact that almost all fields must possess a controlled vocabulary. On some fields, a thesaurus is seen as an essential feature.

In Italy, a lot of work is in progress in this area (see for example Montevercchi and Vasco Rocca (1989)). However, controlled vocabularies have been defined completely only for a few, simple fields. It is worthwhile to note that in these cases, the two vocabularies resulted to be very similar.

An interesting case is constituted by the technical data (material and technique). The Italian approach puts the two items together, as it is considered sometimes difficult to maintain the two concepts separated. French standards, instead, for paintings distinguish between support material and technique, while for other kinds of objects maintain two linked different vocabularies, grouping techniques applicable to the same material.

As far as the artists' name is concerned, even if in both cases reference is made to the Thieme-Becker, the very well known problems (spelling, original name, pseudonyms, school, etc.) may give rise to inconsistencies.

A similar problem is posed by the geographical names (especially if historical aspect is considered too). A possible solution is suggested in Papaldo and Signore (1989), Signore and Bartoli (1989), Signore and Bartoli (1990a), Signore and Bartoli (1990b).

### ***B - The iconography***

In both cases, a great importance is given to the iconographic description. Under this aspect, the very big difference is constituted by the classification system: French standards use the iconographic thesaurus developed by François Garnier, while Italian standards adopt the ICONCLASS classification system. To be more precise, Italian standards recommend the usage of ICONCLASS for the identification of the subject, but identify two different fields: the identification of the subject (where ICONCLASS is used) and the description of the subject (where the cataloguer may insert terms for which an ICONCLASS code is lacking).

It is evident that the reconciliation of the two approaches requires the identification of the equivalences between the two representations in the two different systems.

Perhaps, an agreement may quite easily be found on very general iconographic categories.

## **A - The data exchange problem**

The exchange of data requires the definition of standards at three different levels: conceptual, logical, physical.

The physical level is the simplest to be agreed upon: almost everybody agrees to adopt a 9-track ASCII tape, or a 360Kb MS-DOS diskette.

The logical level is simply the definition of a key for decoding the information contained in the physical support. The exchange format, in this context, may vary from the very simple "card image format" to some sophisticated format, like the MARC format.

Effective data exchange is possible only if a standard has been defined at the conceptual level: that is, when a model has been defined that allows everyone to share the knowledge of the world of interest.

Agreement at the conceptual level permits to understand the semantics of the fields, and may be seen as the fundamental step toward data integration, as is pointed out in Signore (1984), while agreement at the logical level only permits to distinguish one field from another.

It is easily seen that, once the conceptual data model has been defined and agreed, the process of defining an exchange format is straightforward, and may be accomplished overnight.

Even if the definition of a data model may be seen as the most relevant step toward the definition of an exchange format, peculiarities of art history data must be taken into account. As it has been already pointed out, a characteristic of this kind of data is their fuzziness, as the same concept (the name of an object, place, artist) may be designated in different ways, depending on the cultural background of the scholar. The problem is much more complicated than in a conventional business environment, as the normalization of data involves a great cultural effort, in order to reduce to a common frame different ways of thinking, each one based on valid and well established cultural traditions.

## A - Conclusions

In automated treatment of the cataloguing data, a big effort is required to precisely define a cataloguing model, and the structure of the fields.

The italian and french standards appear very close, and under some respects the intellectual approach is very similar. This is especially true for the importance taken by the normalization of the vocabulary and the usage of thesauri and iconographic classification systems.

Some differences are detectable in the general philosophy of the approach.

While the french schema is biased by the characteristics of the target software, the italian standards intend to produce a classification schema that is totally independent on the software and hardware environment, as it is customary in the database design area.

The compatibility of the two standards has been proved possible, but some effort is required by both partners.

The more relevant differences may be detected under some "cultural" aspects, like iconographic description.

## A - Acknowledgements

All the work pertaining to the possibility of the definition of a common standard between France and Italy, has been strongly encouraged by O. Ferrari (ICCD, Italy) and M. Aubert (Ministère de la Culture, France). A great contribution to the in depth analysis of the technical details was given by S. Papaldo and M. Lattanzi (ICCD), and S. Leroy-Beaulieu (Ministère de la Culture, France).

## A - References

- Chen76                      Chen P.P.: *The Entity-Relationship Model: Toward a Unified View of Data*, ACM Transaction on Database Systems, Vol. 1, N. 1, (1976), pp. 9-36
- D'Amadio89                D'Amadio M., Simeoni P.E.: *Strutturazione dei dati delle schede di Catalogo.Oggetti di interesse demo-antropologico*, Istituto Centrale per il Catalogo e la Documentazione (Roma) - Museo Nazionale delle

- Arti e Tradizioni Popolari (Roma), Tipografia Città Nuova della P.A.M.O.M., Roma (1985)
- Massari88 Massari S., Prosperi Valenti Rodinò S., Papaldo S., Signore O.: *Strutturazione dei dati delle schede di catalogo - Beni mobili storico-artistici: Stampe*, Istituto Centrale per il Catalogo e la Documentazione (Roma) - Istituto Nazionale per la grafica (Roma) - Istituto CNUCE (Pisa), 1988 Scuola Tipografica S. Pio X (Distributed by Istituto Centrale per il Catalogo e la Documentazione)
- Montevecchi89 Montevecchi B., Vasco Rocca S.: *Metodologie di classificazione: Suppellettile ecclesiastica*, Centro Di (Firenze,1989), pp. 133 ISBN 88-7038-164-1
- Papaldo86 Papaldo S.,Ruggeri Giove M.,Gagliardi R.,Matteucci D.R., Romano G.A.,Signore O.: *Strutturazione dei dati delle schede di Catalogo: beni mobili*, Atti del Convegno sull' Automazione dei Dati del Catalogo dei Beni Culturali, Roma, 18-20 giugno 1985 (pag. 39-42) Ministero per i Beni Culturali e Ambientali - Istituto Centrale per il Catalogo e la Documentazione (Roma, 1986)
- Papaldo88 Papaldo S., Ruggeri Giove M., Gagliardi R., Matteucci D.R., Romano G.A., Signore O.: *Proposta di strutturazione dei dati del catalogo: Beni mobili archeologici e storico-artistici. (Edizione riveduta e aggiornata)*, Istituto Centrale per il Catalogo e la Documentazione (Roma) - Istituto CNUCE (Pisa), Multigrafica Editrice, Roma (1988)
- Papaldo89 Papaldo S., Signore O.: *Un approccio metodologico per la realizzazione di una banca dati storico-geografica (A methodological approach to producing a historical/geographical databank)*, Multigrafica Editrice, Roma (1989), pp. 573 ISBN 88-7597-105-6
- Parise88 Parise Badoni F., Ruggeri M.: *Strutturazione dei dati delle schede di catalogo: Beni archeologici immobili e territoriali*, Istituto Centrale per il Catalogo e la Documentazione (Roma) - Istituto CNUCE (Pisa), 1988 Scuola Tipografica S. Pio X (Distributed by Istituto Centrale per il Catalogo e la Documentazione)
- Signore84 Signore O.: *Data integration in Art History Information processing*, Conference Automatic Processing of art History Data and Documents, Pisa, Scuola Normale Superiore, September 24-27, 1984, Proceedings, pp. 312-319. Edited by L. Corti
- Signore85 Signore O.: *Architettura di sistemi per la gestione di dati catalografici*, Atti del Convegno sull' Automazione dei Dati del

- Catalogo dei Beni Culturali, Roma, 18-20 giugno 1985 (pag. 51-58), Ministero per i Beni Culturali e Ambientali - Istituto Centrale per il Catalogo e la Documentazione (Roma, 1986)
- Signore86 Signore O., Pardini S., Trasacco A.: *L'automazione del Catalogo dei Beni Culturali: realizzazione della base di dati per le schede OA e RA*, Internal report CNUCE C86-10
- Signore89 Signore O., Bartoli R.: *Managing Art History Fuzzy dates: An Application in Historico-Geographical Authority*, Historical Social Research, Vol. 14, n. 3, (1989) pp.98-104 (Special Issue - Computer Applications in the Historical Sciences: Selected Contributions to the Cologne Computer Conference 1988)
- Signore90a Signore O., Bartoli R.: *A case study for historico-geographical authority*, in Terminology for Museums, Proceedings of an International Conference held in Cambridge, England, 21-24 September 1988, Maney and Son Limited (1990), ISBN 0-905963-62-8, pp.150-158
- Signore90b Signore O., Bartoli R.: *Implementation of a historical/geographical database with support of imprecise dates*, Database and Expert Systems Application, Proceedings of the International Conference in Vienna, Austria, 1990, (Tjoa A.M., Wagner R., Eds.) Springer Verlag Wien New York, ISBN 3-211-82234-8, pp. 271-274