Cataloguing art objects:

a comparison between French and Italian standards

Oreste Signore

CNUCE - CNR
via S. Maria, 36 56126 Pisa

MDA 1991 Conference

European museum documentation strategies and standards

3-6 September 1991

Canterbury, England

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)**
  - **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),**
  - **S. Leroy-Beaulieu (Ministère de la Culture, France).**

Contents

- **Data structuring**
- **The Italian catalog approach:**
  - **History**
  - **The "object"**
  - **The approach**
- **The Inventaire approach**
- **The two approaches compared:**
  - **The data structures**
  - **The vocabularies**
  - **The iconography**
- **The data exchange problem:**
  - **The three levels of standard**
  - **The normalization of the language**
- **Conclusions**

Background

- **Structuring of art history data:**
  - **is not an easy task**
  - **is necessary in avoiding "false drops".**

- **Information must be subdivided into small, semantically well defined chunks.**

- **Too many specialized data structures?**
  Group objects of the same type

- **Identify with a specific tag each element of information.**

- **Structuring is preliminary to the selection of the software and of the hardware**

- **No project successfull if an accurate and clear data structure is missing.**

- **Hypertext and Information retrieval approach need data structuring, too.**

Italian Catalog data structuring: historical background

- **Typewritten cards**
- **General rules**
- **Valuable intellectual work**
- **The relevant issues:**
  - **reduced set of object types**
  - **reduced set of cards**
  - **grouping of information in very general categories**
  - **topological arrangement of cards**

... but ...

  - **conceived for human usage**
  - **semantic ambiguities**
  - **possible inconsistencies**
  - **filled by scholars for scholars**

- **Mid '70:**
  direct transposition in Information Retrieval Systems disappointing.

- **A complete rethinking seemed necessary**

Italian Catalog data structuring: the "object"

- **The object plays a central role**
- **Many object types?**
  Too complex
- **Three diffferent kinds of objects:**
  - simple object
    all his attributes are pertinent to the whole object, and no components which may themselves be considered cataloguing objects may be identified
  - complex object
    - **a simple object whose parts, physically or conceptually separable, exhibit some interesting peculiarities as cataloguing objects**
    - **a set of objects which may be referred by a specific name**
  - aggregation of objects
    when several objects are correlated on the basis of some conceptual criterion, but no name exists which identifies the aggregate
- **Each object is composed by objects.**
- **A specific object belongs to the different categories only on the basis of the quantity and the type of information: no list exists.**
- **It is only a classification model:**
  - **type of relationships between the objects**
  - **criteria of inheritance of the properties.**

Italian Catalog data structuring: the approach

- **Based on conceptual database design methodologies**
- **Entity-Relationship model:**
  - **identification of the basic entities**
  - **identification of the relationships**
  - **no "bill of material" (fuzziness, uncertain data)**
  - **normalization of the vocabularies and authority files**
- **Entity-Relationship model very expressive, but unfamiliar to the scholars.**
- **Representation of the schema via a conventional "cataloguing card":**
  - **information subdivided into small, semantically well defined, chunks**
  - **chunks may be either a field, or 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.**
- **Paragraphs anf fields maps entities and attributes**
- **The resulting structure may be seen as the "linearization" of a non linear text.**
- **Consistency between different cultural areas**
  (semantically equivalent fields are identified by the same tag)

The approach of the Inventaire

- **Methodological choises made in 1964**
- **Based on information retrieval techniques**
- **Massively use of the computers since 1977**
- **All the informations are arranged into fields, grouped into general categories:**
  geographical data and localization, object, chronology, cultural definition, technical data, analytical data, administrative data, etc.
- **Looking at Joconde:**
  - **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**
- **Cataloguing schema conceived having in mind the facilities offered by the adopted software (MISTRAL):**
  (extensive use of the peculiarities of the system, like the possibility of inserting special delimiters in the fields)
- **Each field may be of one of the two types:**
  - **having a controlled vocabulary (or thesaurus)**
  - **free text**
- **From the usage point of view:**
  - **inverted indexes are built on vocabulary controlled fields and on some free text fields**
  - **on the non indexed free text fields only a "secondary" or sequential search is possible**
  - **a secondary search is allowed only on the subset of the database selected by a previous query based on inverted fields.**

The two approaches compared: generalities

- **At first glance the two approaches do not differ very much**

- **In both cases:**
  - **data are presented as a conventional *"cataloguing card"***
  - **information is organized into small fields, grouped toghether**

- **The definition of an exchange format seems to be reduced to the identification of the correspondences between the different fields.**

- **However, there are some basic differences:**
  - **from the "intellectual" point of view**
  - **from the "technical" point of view.**

The two approaches compared: the data structures

- **The italian approach defines a cataloguing schema which is <u>independent on the software and hardware environment</u>**

- **The french schema has been greatly influenced by the underlying software.**
  (But the french standards were initially defined when the database design methodologies were just at the borning stage)

- **The italian schema is much more structured:**
  - **the adoption of delimiters at the data entry stage resulted in some inconsistencies**
  - **small chuncks of information may easily be put togheter, but not vice-versa!**

- **In the italian model:**
  - **emphasis on the relationships between objects:**
    - ***simple***
    - ***complex***
    - ***aggregation***

  - **A component of a complex object is 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.**

  - **Two different types of relationships between objects:**
    - **"vertical" relationship (complex objects)**
    - **"horizontal" relationship (aggregates).**

- **In the french model, the aggregate object concept is missing.**

The two approaches compared: the vocabularies

- **There is a complete agreement on the fact that almost all the 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:**

  - **Production of "Dizionari terminologici" by ICCD**
  - **Controlled vocabularies completely defined only for a few, simple fields.**
  - **In these cases, the two vocabularies resulted to be very similar.**

- **An interesting case: the technical data (material and technique).**

  - **The italian approach puts the two items togheter (sometimes it is difficult to separate the two concepts)**

  - **French standards:**
    - **for paintings distinguish between support material and technique**
    - **for other kinds of objects maintain two linked different vocabularies, grouping techniques applicable to the same material.**

- **Artists' name:**
  - **in both cases reference is made to the Thieme-Becker**
  - **the very well known problems (spelling, original name, pseudonyms, school, etc.) may give raise to inconsistencies.**

- **A similar problem is posed by the geographical names (especially if historical aspect is considered too).**

The two approaches compared: 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:**

  - **<u>french</u> standards use the iconographic thesaurus developed by François Garnier**

  - **<u>italian</u> standards adopt the ICONCLASS classification system.**
    More precisely, two different fields are used:
    - **the identification of the subject (where ICONCLASS is used)**
    - **the description of the subject (where the cataloguer may insert terms for which an ICONCLASS code is lacking).**

- **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.**

The data exchange: the levels of standards

- **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:**
  - **9-track ASCII tape, or a 360Kb MS-DOS diskette**
  - **the same as agree to exchange information by means of paper sheets, or telephone**
- **The logical level is a key for decoding the information contained in the physical support**
  - **from "card image format"**
  - **to sophisticated MARC**
  - **only permits to distinguish one field from another.**
  - **the same as exchanging information on typewritten paper: will a chinese be able to understand arabic or english words?**
- **Effective data exchange is possible only if a standard has been defined at the conceptual level:**
  - **everyone is able to share the knowledge of the world of interest.**
  - **everyone may understand the semantics of the fields**
  - **the process of defining an exchange format is straightforward, and may be accomplished overnight**

Conclusions

- **In the 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:**
  - **importance taken by the normalization of the vocabulary**
  - **usage of thesauri and iconographic classification systems.**

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

- **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.**