

**From data structuring  
to data exchange:  
a simple path**

**Oreste Signore**

*CNUCE - CNR  
via S. Maria, 36 56126 Pisa*

**AHC'91**

**Association for History and Computing  
6th International Conference**

**August 28-30, 1991**

***Odense, Denmark***

# Contents

- **Data structuring**
- **The Italian catalog approach:**
  - History
  - The "object"
  - The approach
- **The data exchange problem:**
  - The three levels of standard
  - The normalization of the language
- **The italian experience**
- **Conclusions**

# Why data structuring

- **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 successful 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 different 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 and 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 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 read arabic or english?
- **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.

# **The data exchange: the normalization of the language**

- **The definition of a data model is the most relevant step toward the definition of an exchange format**  
... but ...  
**peculiarities of art history data must be taken into account:**
  - **fuzziness**
  - **the same concept (the name of an object, place, artist) may be designated in different ways, depending on the cultural background of the scholar.**
- **We are not in a conventional business environment!**
- **The normalization of data involves a great cultural effort (each way of thinking is based on valid and well established cultural traditions)**
- **A lesson taken from the experience:**
  - **effective retrieval of stored data requires a controlled vocabulary**
- **Even better:**
  - **arrange the concepts**
  - **explicit synonymy, preference and hierarchical relationships**
  - **build a thesaurus**
- **The building of a thesaurus is a very long and costly issue. It may be regarded as a long term target.**
- **The normalization of the vocabulary is a task which is imperative and cannot be deferred.**



# The Italian experience: the "giacimenti culturali"

- A Lit. 600.000.000.000 initiative funded by the Italian government in 1986
- Principal aims:
  - employment of young people
  - application of new technologies in the field of cultural heritage management.
- The initiative took the name of "giacimenti culturali" (cultural heritage was assimilated to oil or coal)
  - 39 projects were approved and financed
  - 31 were concerned with the cataloguing of works of art
  - Technological environment was totally arbitrary
  - The only constraint:  
*the results of the projects should be available to the central administration.*
- A general estimate:
  - 15-20% of the budget is invested into hardware and software
  - 80-85% of the budget goes into the education of the personnel involved and gathering of data
- A driving role was played by the classification model defined by the ICCD:
  - standard at the conceptual level
  - included as a constraint in the contracts signed by the firms which were executing the projects

# The italian experience: the "exchange format"

- **The conceptual standard defined in the contract**
- **Logical level standard:**
  - **Adoption of sophisticated exchange format unfeasible**
  - **"card image exchange format"**
  - **defined in a couple of hours!**
  - **A similar approach was followed for:**
    - **thesauri**
    - **authority files**
- **Physical level standard:**
  - **3.5" IBM compatible floppy (360Kb, or 720Kb, or 1.4 Mb)**
  - **5.25" IBM compatible floppy (preferably formatted at 360Kb)**
  - **9 tracks, EBCDIC or ASCII, tape, accordingly to the international standards (1600 bpi, 1/2")**
  - **Names of the files standardized (content and originator)**
- **Everybody was able to conform to these standards**

## Some examples

1	7	80
PV:		
PVU:	piano terra	
PVM:		
PVMG:	lastricato	
PVMS:	irregolare	
...	...	
PV:		
PVU:	primo piano corpo centrale	
PVM:		
PVMG:	alla veneziana	
PVMS:	floreale	
...	...	
PV:		
PVU:	primo piano ala destra	
PVM:		
PVMG:	maiolicato	
PVMS:	a spina	
...	...	

*Example of a repeating group of fields*

## Some examples (cont.)

1	7	80
...	...	
AUT:		
AUTN:	Venusti Marcello	
AUTA:	1512 - 1515/1579	
AUTS:	attr.	
AUTR:	disegnatore	
AUT:		
AUTN:	Buonarroti Michelangelo	
AUTA:	1475 / 1564	
AUTR:	inventore	
...	...	

*Example of a repeating field*

1	7	80
...	....	
FTA:		
FTAN:	ICCD E32336	
FTAN:	ICCD E32447	
FTAN:	ICCD E32448	
...	...	

*Example of a repeating subfield*

## Some examples (cont.)

1	7		80
		OGTD: CROCE DA TAVOLO	
		OGTD: RELIQUIARIO A BRACCIO	
		*SN * Tipologia di reliquiario che rappresenta una evoluzione del reliquiario a mano	
		OGTD: STENDARDO PROCESSIONALE	
...	...		

*Example of a dictionary file*

1	3	7		80
			\$THS	OGTD
			LT	ARTI
				NT ARCHITETTURA
				ARTI APPLICATE
				ARTI E TRADIZIONI POPOLARI
			LT	BACINELLA
				EQ BACILE
			LT	ARREDO PER CELEBRAZIONE EUCARISTICA
				NT CANDELABRO
				CANDELIERE D' ALTARE
				CARTAGLORIA
				...
				RT OGGETTI LITURGICI - PER ILLUMINAZIONE
			LT	...
			\$THS	MTC
			LT	TEMPERA
				NT TEMPERA SU TAVOLA
				...

*Example of a thesaurus file*



# Conclusions

- **Automated treatment of the cataloguing data requires a big effort in order to precisely define a cataloguing model, and the structure of the fields.**
- **A great help may come from the database design methodologies.**
- **Data exchange requires the definition of:**
  - **conceptual**
  - **logical**
  - **physical standards.**
- **A clean data model leads, without any effort, to the definition of data exchange standards.**
- **In a real case, where there was the potential risk of wasting a lot of resources, starting totally uncoordinated cataloguing projects, the availability of a cataloguing model resulted in an effective discipline of data.**

# Acknowledgements

- **O. Ferrari,**  
at the time being the director of the Istituto Centrale per il Catalogo e la Documentazione (Rome), who pursued the cultural objective of building a "Catalog" and not simply making an inventory (or counting the objects)
- **S. Papaldo,**  
who coordinated the activities of the ICCD in defining the standards and gave a great contribution to the definition of the reference model for art objects
- **M. Ruggeri, L. Cavagnaro, and others from ICCD,**  
who performed an in depth examination of the cataloguing data, testing the approach on several case studies.
- **My colleagues:**  
**R. Bartoli, R. Gagliardi, R.D. Matteucci and G.A. Romano** for their help in the definition of the general schema and support in several case studies.