

Manipulating information, Producing data

*Author:
Ifigenia Dionissiadou*

CIDOC06
GOTHENBURG
S W E D E N

Within the last two decades the computerized documentation systems available for museums have undergone a lot of progress. The most well-known among them follow the international standards and have been well established in the museum market. Even though the standards regarding the information categories and the ideal structure of a typical museum system have been studied, analyzed and developed in a satisfactory degree, issues regarding the content of the information still provide distress to the documentalists and the museum users and need to be further developed.

The primary difficulties regarding the handling of content in a museum documentation system can be separated into three categories: the *selection* of the appropriate content to be entered, the *systematization* of this content and the *structuring* of the information provided in this content in several levels of retrieval.

In most cases the computerized system replaces the previous manual methods of documentation, consisting of many hand-written sources: inventories, catalogue records, indexes, publications and other paper archives. These resources often vary, depending on the age, the size and the economic wealth of the institution. The diversity of quantity and quality of available hand-written sources leads to important difficulties to the procedure of selecting the type and the amount of information to be input in the computerized documentation system. It also raises questions regarding the documentation of the identity of each piece of information, that is registered and the important role of metadata, concerning the author and the source of each established information unit.

A second consideration for the selection of the content arises from inconsistencies between different versions of the same information, with which the documentalist is often confronted, when registering information from several different sources. One of the major advantages of computerized documentation is that it provides a very effective tool for the juxtaposition and the comparison of the several different information sources, allowing the museum to gain an almost absolute control of its information content. Alternative suggestions about the dating of the object, its assignment to a specific artist, attribution of its origin to a place, as well as information regarding its type and usage, have often evolved as different sources were created. These alternative and conflicting suggestions lead to confusion for the documentalist, who, in most cases, is forced by the system to choose the most accurate and disregard the rest. It puts an unreasonable burden of responsibility on the documentalist, since it is very difficult to establish

specific rules regarding such decisions. Even in the rare cases whereas the curator and the documentalist are the same person, and a responsible decision is taken, issues remain, concerning the documentation of this decision-making process within the system: other users of the system may come upon discrepancies, between data recorded in the system and in the various previous sources, which may lead to a loss of confidence in the content contained within the documentation system, or to a misguided "correction".

Variations between hand-written sources are also often due to the inevitable mistakes that happened in the course of the several recopying procedures, which took place in the past. This is the most common reason for misspellings of peoples' names. It is also sometimes the reason for misleading assignment of ID numbers to museum objects. In most such cases, the oldest source of information proves to be the most accurate. However, nothing can be so troublesome for the documentalist as the existence of a problematic numbering system of the museum objects, resulting from partially executed, but unfinished, attempts to revise the numbering method. This old mismanagement can lead to double and triple assignments of the same inventory number to irrelevant museum objects and takes large amount of time and energy, by many people, in order to be permanently resolved.

In retrospective documentation, selecting the proper information to become the content of the museum documentation system is, in conclusion, quite a demanding and time-consuming task. Retrospective documentation often raises questions regarding the management of museum resources, but may often contribute to the resolution of chronic documentation problems and misinterpretations. Retrospective documentation is, in effect, an extensive research project, currently underestimated, which requires the support of the museum curators. The extent of the research into existing direct and indirect museum sources provides the richness and the *maturity* of the computerized content.

The systematization of the information about the museum objects is the second issue to be discussed. Controlled terminology is used by most of the computerized documentation systems, for defining the most essential information categories of the museum objects. Pick lists of terms and the classification of controlled vocabulary in a thesaurus structure are the most commonly available tools, which facilitate data entry and guarantee accurate information retrieval, when questions based on multiple criteria are posed. Nevertheless, the imposition of controlled terminology, even though understandable, is regarded as a controversial methodology by a large number of museum people. A major reason for this resistance is that, with the exception of the

managerial information, most of the information regarding the *description* of museum objects cannot be classified in a systematic way without sacrificing some of its essence. For the curators, who bear the ultimate responsibility for the description of their collections, the use of variations in terminology, expresses fine differences between meanings, important, since these differences embody much of the cultural richness of the collection.

The process of producing controlled terminology becomes progressively more controversial as a function, by the increase of the level of detail in the descriptions and the extent of the variations among the groups of objects that are handled as part of common retrieval. Additionally, conflicts of terminology frequently arise from shifts in meaning, that specific terms may undergo, according to the cultural context in which they are used. The meaning of certain terms may also change according to the group of objects they describe. In general, it can be argued that the systematization of cultural information through controlled terminology has a negative impact on precision of expression and imposes tough rules in a delicate context. In addition, the amount of metadata needed for the user and the documentalist, in order to be able to use controlled terminologies accurately, is disappointing and sometimes even frustrating: nobody, even within the same museum can easily describe the features that differentiate a bowl from a cup, or a dish, especially when these terms are describing objects of unknown usage. The absence of international rules and definitions creates an important, but understandable gap.

Understandable, because artistic and historical objects are unique by choice, for this reason such objects tend to have only limited uniform features. In most cases they were chosen for their rarity and their peculiarity. These exceptional characteristics are important aspects in their description. Additionally, artistic and historical objects originate from successive layers of cultural periods, some of which emerged in a sequence, whereas some others evolved simultaneously in the same or in other places. It is inevitable that the terminology used to describe cultural objects has a partly historical character. In this sense, it cannot be replaced or modified without losing its precision. Historical information, being multi-dimensional, representing complex relations in time and space, cannot readily be reduced to simple parent-child hierarchies. A better methodology needs to be developed in order to make such classifications effective and usable; or to achieve the same goals of ease of retrieval without the need to sacrifice variations in terminology and the subtle expressive power of natural language.

Cultural information also has another much neglected, but nonetheless important characteristic: it might be inaccessible, or partly accessible, and still keep its high value. Historical and

archaeological research only partly recovers the past. Unknown and half-known elements are always worthwhile for further consideration, so interrogative and disjunctive expressions need to be placed as information qualifiers within the museum's documentation.

The considerations discussed above, regarding the systematization of important cultural information of the museum collections, could perhaps be eliminated if a clear separation could be made between “essential” information units and their systematized form, which would be stored separately and used only for the purposes of retrieval. In this case, the demanding eye of the end user would not be disturbed by the presence of uncommon, repetitious and imprecise controlled expressions. The documentalist, on the other hand, would be free to use in this hidden *meta-record*, for as much encoded terminology as would be needed to ensure the granularity of the retrieval, without the need to be too concerned about linguistic considerations.

The final issue I wish to discuss is the need to structure information in units, in order to allow retrieval at multiple levels of specialization. Cultural information is not only expressed in various levels of precision, but it is also retrieved in several different levels, according to the users' needs. Questions regarding time and place may be either expressed in common values of centuries, years and dates, continents, existing countries and cities, but may be also expressed in historical-cultural values of artistic or ruling periods. Questions regarding object name, material and technique often need to use vaguer terms, in order to widen the scope of the retrieval. For the above reasons, museum documentation systems need to include time primitive and classification tools that allow the construction and the usage of multiple thesauruses.

The computerized documentation system, being a very expensive and extremely time consuming investment, is implemented with the aim of covering the needs of all the museums' different users: the curators who want to incorporate all the managerial and scientific information about the object, as well as their personal notes regarding an ongoing research; the conservators who need to be able to input detailed information about the state of the object and to describe its conservation treatment; the researchers who wish to interrogate the database and to be presented with all the available historical and cultural information, expressed to the highest-precision and using all scientific means; the museum visitors and the virtual visitors of the web, who need a basic summary in more than one language; and the students and children, who are interested in the information which has educational and possibly entertaining features.

These high expectations of information presentation and retrieval are difficult to cover with just one system. However the inclusion of information qualifiers in a documentation system may help to fulfill some of these requirements, as it allows the inclusion of related terms, synonyms and translations. The limited implementation of documentation systems that are able to use multiple hierarchical thesaurus structures is an important obstacle to a more powerful categorization, which would lead to better presentation and retrieval.

In the end, I would like to express the opinion that there is much more to be done in both the theoretical and the practical aspects regarding cultural documentation systems. There are still many open issues which need to be resolved if progress is to be made towards the implementation of better operating solutions: solutions that are less complicated to use, require fewer compromises, and are more accessible to less specialized clients.