

Tomasz Zaucha

National Museum Kraków

Tarde venientibus ossa? Museums' data today and digitization strategies

Museums are institutions that last. In fact, they emerged fairly recently, in the modern period, and only few of them are older than two hundred years. Nevertheless, museums are intended to last, it is hoped, eternally, and their founding charters and legal environment stipulate provisions that should make them immune from the fortunes and misfortunes affecting individual people, or even countries. They should be as stable as a rock in the middle of the stream, and, like old shrines and cathedrals, they should shelter and preserve for future generations the most precious pieces of the past and of today. This view is not only a legacy of romantic museology, but still remains a mind frame for museum professionals today. It very often stands in striking contrast with the course of modern life. The development of technology, especially computing and communication, creates one of the areas where the tensions between the permanence of museum and fluctuation of today's ideas come to the fore.

Since one of the primary objectives of a museum is education, which comprises both the description and interpretation of the collection, one of museum's main functions is to create and develop information. Here the paths of museology intersect with the fervent world of information technologies. The benefits of the latter are obvious, especially in collection management and object identification procedures where the ease of creating visual records and their multiplication is, I dare say, revolutionary.

Museums therefore have been trying to embrace the new tools and use the new possibilities. Gradually, IT departments became standard units within a museum's structure, but apart from putting computers on desktop and replacing pens with keyboards, they hardly solved the new, specific problems that appeared on the horizon. One of the most obvious ones was to replace the traditional museum inventory with a computerized one, but it soon turned out that the extent of different problems and solutions was much wider than anyone could have expected. There appeared a multitude of solutions, theories, proposals and trials, instruments and applications, many discussed in-depth also on this conference. Gradually, not only electronic documentation but rather museum information technology has become an important branch of museology, and, I hope I am not exaggerating, also of scholarly information in general. Nevertheless, at least on the Polish ground, the information science theory and methodology did not take root deep enough in the field of museology. The existing literature on specific problems that museums cope with right now: e.g. databases, data standards or thesauri building, is indeed available in Polish, but was written by people specialising in library science and bibliography, and is concerned exclusively with the problems encountered in libraries. Whereas the theory of information, bibliography, collection management, and computer science are well established topics in library science departments in many universities, the same areas, but applied to museum collections (or art and cultural heritage in general, like heritage protection state service), did not enter university syllabi at art history courses until very recently.

The other problem that affected the situation in Poland was the absence of a coordinating hub that could serve as a platform for at least an exchange of problems and solutions, for creating leader groups or even lobbies, and, eventually, maybe for forging out standard solutions and tools that would be ready to apply by any and all museums in the country. Here, again, the difference between library science and museology comes to the fore, because in Poland the National Library in Warsaw has been the coordination center for most state libraries. It introduced metadata standards like MARC BN (suffix BN stands for Biblioteka Narodowa – National Library), which was a national Polish variant of the American standard (now being widely replaced by MARC 21, again – under the auspices of the National Library), and maintains the subject headings department which employs a reasonable number of personnel and functions as a country-wide validator for the headings. The other such center is the library of the University of Warsaw, where the second

widespread subject heading authority is created (KABA, based on the French RAMEAU).

The efforts of libraries stand in stark contrast to those of the museums that tried to cope with such problems, but usually did it on their own; a group of three museums on the Baltic coast that created their common thesaurus is a singular exception to the rule. Moreover, there are only few institutions that tried to keep pace with the current of museum information science, not only by observing the new developments, but sometimes possibly also applying them. A handful of museums have tried to accommodate to the internationally recognized metadata schemas, like Spectrum, and Categories for Description for Works of Art. There have been three attempts at creating unified thesauri for collection description and categorization (the above-mentioned museums of the Baltic region: National Museum in Gdansk, National Maritime Museum and Castle in Malbork (Marienburg) (2013)¹, National Museum in Krakow (2000), and Art History Institute of the University of Wroclaw² – the last of the thesauri was not applied in any museum). Then, (since 2009) there has been a growing number of museums that present their collections on-line (the hitherto greatest collection exceeds 90,000 items, accompanied by more than 40,000 digital photographs). Unfortunately some of these projects aim rather at creating playful and impressive portals that employ animation, virtual visiting, costly 3D models, and the like, than at extensive digitization of the museums' vast assets. Nevertheless, the museums that had planned to enter the digital world, as a rule, simply used to buy (and still do) one of a few (usually one of two) softwares written in Polish – essentially created in the 1990s – and sometimes customize them as they seem convenient for their own purposes, and their employees start to type, creating all data from scrap. Such a database is then used mainly for creating a digital mirror of the analogous documentation, and, consequently, for the collection management: storage, loans, internal queries, digital photography archives.

Multiple reasons have contributed to this state of affairs. Clearly, finances are always a factor. Museums usually have more important expenditures than state-of-the-art computer equipment. With their specialised needs and reduced means, museums do not make a good customer for application development companies; what is more, the market is limited and shallow. And museums can hardly compete with business companies in hiring highly qualified computer specialists.

The human factor, which is often neglected at various enterprises, seems to be one of the most significant here, and what we mean here is not only IT professionals but rather the museum people: the management and curatorial staff: those who make decisions, and those who create data. There is also usually a small, but pivotal group of curators or registrars that have been assigned the role of digitization supervisors or coordinators. Registrars are usually charged with this task because in Polish museums, a registrar is a person who oversees not only the collection management, but the museum documentation in the first place. As for others, they are usually recruited from junior staff, on the simple premise that young people are good at computers in general. That is usually true, and they are committed people, but their placement does not give them much opportunity to influence management and senior curators. All this, paired with a limited possibility of training in the field of museum informatics, results in the fact that the equivalent of a “system librarian” - position that appeared in Polish libraries as early as in the 1990s, is absent from museum terminology – and life.

Some museums, as a rule – big museums – fare better than others, for obvious reasons: they employ better qualified personnel, they are more successful at making contacts (also international ones), participating in important projects, updating hardware and software, and at providing development opportunities (e.g. the acquisition of the most recent literature). Last but not least, their large collections pose more problems, which raises the level of awareness of the extent and diversity of issues that museum informatics is concerned with.

The museum digitization in Poland, as I have presented it here, paints a rather gloomy picture. We are late, and the problem is whether we can adjust our development to what is being currently done at the international level, whether we would be able to create and apply solutions

1 http://www.nmm.pl/upload/Files/cke/Digizaurus_material_1.0.pdf

2 <http://historiasztuki.uni.wroc.pl/tezaurus.html>

equivalent to those we know from elsewhere, at least such basic ones as a standardized metadata sets. Yet the picture is not altogether dark. Decent solutions appear here and there. Moreover, not long ago the Ministry of Culture and Heritage established the National Museology Institute in Warsaw, an institution that— among other things – acts as a competence center for the digitization in museums. The institute is a body that, hopefully, will be able to coordinate the efforts of museums, and, what seems vital, it has already developed a training program for museum professionals, aimed both at theoretical and practical issues in museum informatics, and, specifically, the digitization of assets. The appearance of museum informatics at university courses, and the growing consciousness among the management and senior staff are good perspectives for a change in due time.

But still, the impression of delay remains when we look at the data and compare it what has been done, and what should be done. It is estimated that ca 120 museums in the Lesser Poland region possess roughly 4 million items. The National Museum in Krakow (together with the Czartoryski collection) which is the greatest in the region administers ca 800,000, of which 380,000 are in the system, and 90,000 have been published on-line³. I have not been able to acquire information how much data the other museums in the region have in their databases, but the estimated number of their holdings published on-line certainly does not exceed 100,000⁴. In the light of these numbers it is clear that the more technologically advanced bigger museums, do not create majority of data. The majority of items are dispersed among smaller, or quite little institutions.

If we make an assumption that the main aim of digitization is to give full access to museum collection, then we must admit that, notwithstanding all theoretical issues and software and hardware problems, it is the data input that is the main factor in building long-term digitization strategies. And since digitization means not only – as many still believe – creating digital image of an item, but also (and in my opinion in the first place) creating digital metadata, or simply: a keyed-in description, the latter action must be taken seriously into account. The first and most important factor is time. Creating such a vast amount of data will take years, which in terms of technology development means eons. The applications we use now will soon be history, and the hardware will be long recycled. It is quite probable that within years also metadata standards or resources and structures like thesauri or ontologies will be outdated or replaced by newer and more advanced ones. All this means that the pursuit of the most recent solutions, and following the newest developments must be prudent, and only well-grounded and best established systems should be introduced. This means also that the theory of museum informatics must not prevail over the knowledge of one's discipline, whether art history or archaeology or ethnology etc., and in many cases the experience resulting from research may get a priority. This is good news for museum professionals who do not have much, or even any, experience in the field of digital museum documentation. We may reassure them that they do not have to know much about, e.g. Conceptual Reference Model, in order to be able to create valuable data. The other important factor is that nowadays we are in possession of far more numerous instruments, several satisfactory free applications, and data standardization favors open solutions. There is little risk that data once created will be irretrievable on later stages, or cannot be migrated; the opposite is true: data created even with the simplest methods, e.g. recorded in a table in OpenOffice Writer, can be surely – sooner or later – imported into a collection management system or transformed onto CRM. Surely, museums would rather choose a ready-made application, one that runs collection management; nevertheless it is worthwhile to realize that such proprietary software is usually ill-adapted to data re-use, export or migration (except for the upgrades).

This attitude – data first – seems to be most advisable. Nevertheless, before we encourage curators to start working, some essential caveats must be clearly articulated. Any data, from any stage, will be eventually used in advanced applications, both in collection management systems and in on-line databases, therefore their form must enable further conversions. Whether we set up an

³ www.katalog.muzeum.krakow.pl

⁴ Under auspices of the National Museum in Krakow, the first Polish aggregation catalogue of museum collections was published in 2014: www.kultura.malopolska.pl

interim tool or customize any more advanced application, 'granulation' is the keyword. This simply means that data must be chunked into the smallest possible pieces so that each of them can be processed separately.

The second imperative is consistency on any possible level: personal, departmental, organizational: each piece and type of information should be recorded in the same shape. Fortunately, in case of unavoidable inconsistencies, distortions and errors, the modern search algorithms are a factor of indulgence. The failure to apply appropriate granulation may be usually repaired only by tedious human work.

Keeping the above in mind we would like to repeat that the sound knowledge of one's particular discipline must be applied in any case. In simple terms this means that when we want to create an entry e.g. on stained glass windows, we should follow the *Corpus Vitrearum*, or when old prints are being digitized then the recent editions of Bartsch and of Hollstein will provide exemplary solutions. The way the information is formulated as well as the type and extent of the information provided should be chosen carefully, especially when using predefined structures of museum databases: the fields labels are often false friends. The character of the content should therefore be a subject of deliberation, and of discussion among users. Those should lead eventually to setting cataloguing rules, which is not an easy task, and cannot be done in a short time. Of all the aspects of museum data standardization, it is the data value standards, and not the structural ones, that seem to be more important, but they are more difficult to compose. Whereas in libraries the data value standardization is already highly developed, there is only one museum in Poland that has managed to draw up a cataloguing handbook for its personnel.

It is worth stressing that such rules encompass many types of data, and cover more issues than controlled vocabularies do. Building such a vocabulary, and, especially, building a thesaurus, may be a contestable task of structuring the knowledge. This aspect reveals itself clearly when multilingual instruments are in question, since concepts do not necessarily match. But then the thesauri in the process of day-to-day data input do not serve as knowledge structures, but rather as a simple accumulation of terminology. Fortunately, nowadays there is a number of auxiliary tools; in most cases they need not to be implemented into a museum's system to be used for clarifying terminology and sustain consistency.

The important issue is the source of data. The analogue museum documentation, sometimes reaching many years back, can in many cases be still useful, since in the humanities the knowledge, not only the established facts, but also their interpretations, often retain validity for long periods of time. This may mean that digitization of the collection data may be carried out with the help of typists, not curators. Nevertheless one must be aware of the duality of data, and should discern between identification, with elements like dimensions, technique, and basic photograph playing the leading role, and research information, where e.g. date, attribution, place of origin, are the most important data. Because of the usually great size of collections both types of data cannot be updated constantly. This is a good strategy when we use the data purely for documentary purposes. Anyway, changes in a book inventory cannot be made freely and all the data are regarded correct as long as they enable us to identify the object without doubt. But the usage of the same data for publication purpose may be considered dubious by many. In this view, the instant on-line access to the collection information means access to up-to-date information, which is not necessarily true. Here digitization helps to distinguish between a mere accumulation of facts, and the research proper. In the nearest future, as digital data grow older, there will appear a need to draw necessary distinction lines, that are so clear in print: anyone can tell a simplified inventory from catalogue raisonné or early editions from the newest ones.

When speaking of museum data, and digitization one must not forget of digital visual documentation. In some common concepts of digitization, digital photography or scanning, plays the leading role, and is even essentially regarded as the sole purpose of all action (which is not a correct view). Here the situation is more than simple: although the state-of-the art photographic documentation is most welcome, then any photograph is better than none.

I am perfectly aware, that many aspects of museum assets digitization, that I stressed as important, may retain their validity only for a time. They are, after all, interim solutions. Nevertheless, keeping in mind how the vast resources still have to be converted into digital form, whether by retyping the analogous documentation, or by retrieving new information from the items themselves, the term 'interim' (Latin 'temporary') may mean the time “limited” for long enough, to catch up with any current developments. Whatever the eventual solutions will be, the data building is the task for now. Let's start.