

Evaluation of the survey of the level of digitization in museums in the Czech Republic

David Cigánek, Moravian Museum - CITeM

The continual application and development of digital documentation in museums and galleries is one of the principal pre-requisites of fulfilling their role as memory institutions in a modern information society. Meaningful support of digitization on the part of the Ministry of Culture requires that comprehensive information on the current state of digitization, the needs of the individual institutions and their ideas concerning the way ahead is available.

Survey methodology

The questionnaire survey initiated by the Museum and Gallery Department of the Ministry of Culture of the Czech Republic (Mr. Michal Janiš) was carried out in liaison with the Museum Information Technology Centre (CITeM) in March and April 2006. A total of 289 museum institutions registered in the Czech Republic were called upon to take part in the survey, with a special focus on those which have their collections recorded in the Central Collection Register – from small private or local community museums to medium-sized regional museums and to the largest institutions directly under the administration of the Czech Ministry of Culture. The questionnaires could be filled in by hand and sent by surface mail; in a text editor on a PC and sent via e-mail; or the required data could be entered directly in a web form available on the CITeM server. The fairly high number of the responses collected, as well as engaged comments and notes by respondents, confirm the highly topical nature of this problem.

Survey results

First of all, the survey organizers would like to extend their thanks to all the museums, which filled in the questionnaire in a proper and prompt manner irrespective of the technical and administrative complications which were occasionally bound to occur. Any concerns that the questionnaire survey would be dismissed as meaningless hyperactivity on the part of the staff of the Czech Ministry of Culture, soon proved to be unfounded due to the positive approach by most of the institutions, including municipal and private museums whose management has virtually no relationship to the Czech Ministry of Culture whatsoever.

The survey outcomes commented on below are therefore the result of processing a sample of data from 170 museums throughout the Czech Republic, from the smallest establishments to

the large institutions such as the Moravian Museum and the National Technical Museum. The nearly 60 per cent return rate (58.8%) may be considered a great success given the length of the questionnaires.

Collection Management Tools

Collection management, or the management of other sets of items, is facilitated in most museums by using one of the two large CMS (Collection Management System) environments available on the Czech market - Demus and BACH. In the largest institutions both systems may be used in parallel at different workplaces, often in combination with various types of "home-made" software. The share of the two systems is roughly equal with Demus having a slight edge. However, there is a distinct difference in the amount of stored data, where almost half of the records (47.8%) are processed in the structured databases of Demus. Thanks to the foresight of the Czech Gallery Council it was possible to achieve a state in which the whole gallery community uses the museum CMS (in all cases Demus, with only one exception). This ensures data and methodological compatibility throughout the field, the absence of which elsewhere, for example, in Slovakia, now causes complications.

In principal, nothing stands in the way of general deployment of CMS even in the circumstances of the smallest institutions with a single summary collection. The need to use one's own software tools usually springs from the need to process highly specialized collections not covered by the existing collections modules of Demus or BACH; the specific conditions given by the exclusive position of the institution (National Museum); or, most often, unawareness of or distrust on the part of the staff responsible. They are, in general, simple database applications built around the MS Access platform, modified library software (ISIS, MAKs,

Clavius, KP-Win, etc.) or simple data sheets in MS Excel. In small museums we may even encounter old database/record keeping programs for MS-DOS – such as AISM or applications in Fox Pro, Paradox and FAND. The weak spot of "home-made" or custom-produced applications is often the non-standard structure of the database table row, infrequent use of relational tables (codelists) and, consequently, difficulties in data export/import to/from other systems (incompatibility). Databases of the above type currently store more than 800 thousand records, i.e. approx. 16% of all digitized text records of collection items.

Text digitization

Text data provides descriptive information on the individual collection (or, if necessary, non-collection) items. Given the situation in Czech museums, text digitization actually involves manually entering information from inventory cards and/or additional accompanying information (catalogues, location labels, etc.) in a PC database application.

Most museums are aware of the great demands on work capacity related to text digitization and, at the same time, the limited (human and financial) resources restricting them in turning this activity into a routine. Only about one quarter of the surveyed institutions (25.9%) inputs the basic set of data on the collection items into the database thus endeavouring to fulfil the imaginary quotas and, more importantly, to create, as soon as possible, a fully functional digital database of the collection items in the institutions. The remaining museums (74.1%) opted for a strategy of the most complete description of a collection item in the database in order to avoid repeated returns to once digitized collection items and to eliminate additional handling of items within the collection. This approach naturally takes up more time and makes greater demands on the expertise of the staff making the entries who would be able to precisely analyse many particular field-specific descriptive features of an item.

Presently, over 5 million items have been digitized, which represents an enormous volume of work by collection curators and documentors. With an estimated existence of roughly 65 million collection items in Czech museums [Žalman 2005] this is just a little more than 7% of the total. However, we should bear in mind that

by doing this the accessibility of the data for researchers and the public has greatly improved thanks to the sophisticated options of database processing and the publication potential of the Internet compared to objects kept in dusty vaults and documented only on paper cards. In addition, the text data provides a basis for adding further digital documentation – including the attractive and much demanded image and sound records.

Image digitization

In general, slightly more than a half of the respondent institutions (52.9%) are involved in some form of image digitization. In the group of museums following the way of slow, but more thorough text digitization, we notice a higher proportion of institutions which perform, in some way, image digitization. In organizations processing their collections using a quick, basic description it is only one tenth of all the respondents. Image digitization is evidently considered to be a superstructure over the traditional text description of a collection item, rather than a clear target of museum documentation work.

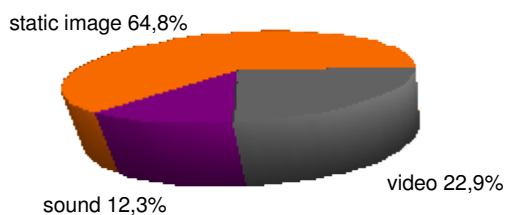
All in all, the questioned institutions store about 300 thousand digital image records with a total volume approaching 7.5 TB of data. The actual volume might be even greater, as the value excludes data-in-progress, i.e. at the stage of acquisition and the next stage of processing at specialized workplaces. Given the fast pace of development of digital imaging technology and its resolution we can shortly expect a sudden growth in the number of data files and their volume in particular.

Sound and video digitization

These two types of digital documents appear in the submitted questionnaires with a much lesser frequency as their use is tied to specific fields of museum work. In both cases it is mainly ethnography and the vast digital audio & video archives of the Technical Museum in Brno. The markedly different ratio of the volume of audio and video data of the two great managers of ethnographic collections is striking: the volume of audio data in NFCI Strážnice is two-and-a-half times greater than the volume of digital video material, while the Valachian Open-air Museum (Rožnov pod Radhoštěm) manages an almost 1 TB video archive in contrast to merely 14 GB

of audio data. Unfortunately, the information obtained from the questionnaire does not allow us to directly determine whether it is a consequence of differing approaches to ethnographic documentation or the method of data storage in different data compression formats. In total, the archives of the respondents hold about 1.4 TB of audio data and 2.6 TB of video data.

Fig. 1 - Proportional distribution of several types of digital data



Technical equipment

The basis of digital documentation in museums consist both of the entry of text information on collection items (see above) into a structured database and image digitization which, in itself, has become a synonym of digitization in general for many museum staff. No wonder then that the most frequently owned item of technical equipment is a digital camera, although it still does not feature on the list of equipment in all the museums. At least one is owned by 85% of institutions. In medium-sized, and especially, large museums it appears in individual departments or even workplaces. Archaic models with a resolution of 1.2 – 2.1 megapixels (MPx) appear only sporadically, given that they are completely unsuitable even for very basic documentation work due to often poor image and colour quality. A much more frequent group is made up of standard medium-class models of various age, and with a corresponding resolution of up to about 5 MPx, fully sufficient for everyday documentation work in the field and a makeshift studio. Institutions which think seriously about digital documentation furnish themselves with professional equipment often designed as a reflex camera with exchangeable lenses and a resolution of up to 13 MPx. These can be used

for the most demanding applications in intensive operations – a pre-requisite is a well-equipped studio and fully-competent camera operators. Most of the respondent museums consider a resolution range of 2-4 MPx acceptable for most documentary activities, and then up to 8 MPx for more demanding applications, which is in line with realistic technical requirements for digital image quality.

A similar percentage of institutions (85%) is equipped with a desk-top scanner for recording flat items. Almost all scanners support a resolution of over 1200 dpi, although even for the most exacting work on digitizing, for example, old engravings, the resolution used hardly exceeds 600 dpi, while for general work the recommended resolution used by most of the museums is 300 dpi. The museums can be considered adequately equipped for standard applications. Problems may arise when capturing large-format or sensitive objects.

Recording a dynamic image requires greater experience and a special approach in further processing (editing). It is by no means a standard method of museum documentation, more widespread, for example, in documenting the folklorist phenomena in ethnography and animal ethology in zoology. Only 29% of the respondents are in possession of at least a simply equipped digital video camera, while more sophisticated models featuring three-chip optics and a higher resolution are less common. Apart from the above described applications, the purchase of a video camera may be also motivated by the possibility of filming events organized by the museum – such as exhibition opening days, and the "Museum Night", or "Open House" events, etc.

In the note on additional equipment some museums mention using digital backs, overhead projectors, DAT recorders and dedicated film scanners.

Digitization procedure

In the approach to the volume of necessary digitization we find significant methodological differences. Some (especially the larger) institutions understand digitization primarily as a tool for providing access to however small section of the collections - for example, through a web presentation - elsewhere the targets set for digitization come close to the complete processing of digital documentation of all

collection items in all the collections.

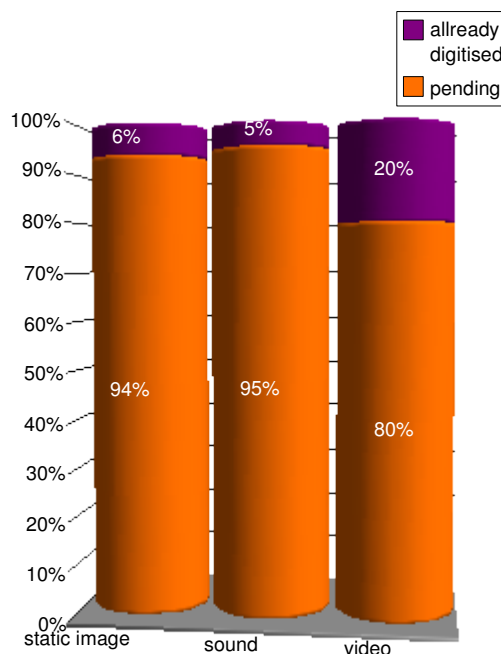
When all the methodological requirements for the digitization procedure are to be observed it is impossible to realize complete digitization of all the collections of an institution within a short-term period. Small museums (private, municipal...) have a clear edge managing only a limited amount of collection items, often deposited in a single building, if not a single repository. In this case, even limited human resources are sufficient to completely digitize the collections within a few years, as practical examples confirm. Also, the fact that the collection manager or documentor can "see" to the end of the whole digitization campaign and does not take it to be a never-ending monotonous activity without any obvious contribution to his personal development is of immense importance. On the contrary, in large museums with hundreds of thousands of items, where simple stock-taking lasts several years, it is physically impossible to speed up digitization so that it would be completely finished during the life-time of the current manager of the gigantic collection. This exerts a strongly demotivating influence and deters custodians both from actively participating in digitization and can even develop into a negative attitude towards working with ICT in general. In similar situations it would seem reasonable to consider establishing a permanent digitizing workplace equipped with top recording technology that will be efficiently used on a sufficiently large volume of new acquisitions. In small institutions with a relatively low number of acquisitions per year the optimum solution seems to be third-party digitization on a collaborative or commercial basis, with financial support by the founder and/or the Czech Ministry of Culture.

The graph makes it clear that museums and galleries are still at the start of a long journey. It should be noted, however, that the total number of items for digitization is by far below the number of items in exhibition halls and repositories. It would not be economical to digitize many of the items due, for example, to the disproportion between the museological value of the object and the cost of digital documentation (fragmentary archaeological finds), or the low indicative value of the digital document (entomological and botanical collections). It can therefore be taken for granted that the collections in our museums and

galleries will probably never be completely digitally processed, at least not down to the level of the individual physical objects.

The existing situation in static image and sound digitization is very similar to the percentage of processed material. More accessible and attractive digital imaging technologies facilitated digitization of about 6% of the planned volume during their existence, while sound archives reached a similar share (5%) also due to the fact that for many sound recordings made by now archaic technologies speeded up digitization is virtually the only realistic procedure to enable their playback and thus continue their sustainability. As far as digital video is concerned, nearly a quarter of digitized documentation can be ascribed to the relative novelty of the technology – when most of the video recordings were made only recently, using contemporary digital video cameras and hence directly in digital format.

Fig. 2 - Digitised / pending rate



Data formats

All the respondent institutions which digitize static images store original data files. The total number of the files is at least 185 thousand – however the estimate is not very accurate as it incorporates only the respondents who were at all able to quantify the great number of image files and did not resort to stating the total volume

in gigabytes. If we extend the calculation by including this data as well plus original image files in institutions who avoided filling in the questionnaire the grand total is very likely to exceed half a million images.

The original files then undergo basic graphic processing to create presentation formats in high quality (used by researchers, in catalogue printing, etc.) – which is performed by 52% of museums – and thumbnail images to enable quick and more practical searching through image sets or to suit the purposes of publishing, for example, on the Internet - 42% of museums. Again, the given number of almost 300 thousand high-quality derived images may not be very accurate as in some institutions there seem to be some confusion between the terms "original data file" and "data of high quality". The number of derivatives will therefore be significantly higher, not to speak of the unidentified number of thumbnails often created in batches or in a semi-automated manner as the needs of their integration into the collection (image) registers require.

The basic data format for storing images when working with a digital camera is, in most cases, JPEG, sometimes in combination with TIFF. Only infrequently do the respondents mention formats such as RAW, PNG, and very rarely Kodak Photo CD, SHQ and BMP(!). In images captured by means of a scanner the most frequently appearing format is BMP, and rarely PSD, both being a consequence of the forced choice dictated by the scanning software. The JPEG format also dominates the domain of working copies and thumbnails where again the TIFF and BMP formats are encountered very infrequently. Regarding the role of museums as memory institutions, the key problem is primarily long-term data storage in sufficient quality and with guaranteed accessibility ensured by recording in a documented and widely supported format. The JPEG format is exclusively used for archiving purposes by 43% of museums enabling them to use suitably low compression settings to achieve economical storage of great volumes of data of reasonable quality taking up a relatively small volume of storage media. A half of the above institutions (i.e. 21%) uses uncompressed or compressed, with no-loss compression, TIFF format for this purpose; 36% work simultaneously with two or three storage formats (most frequently JPEG and TIFF). There

have been museums which utilize formats unsuitable for creating archives - RAW (a proprietary format dependent on a specific digital video camera manufacturer), BMP (very uneconomical usage of disc space) and Kodak Photo CD (not widespread and with uncertain future support). It is hoped that, given their mentioning of other types of formats used in parallel, the less suitable ones are only applied in alternative storage.

Only up to 16% of institutions work with digital audio data. The most frequented format is the well-documented WAV, which stores data without further losses by compression and can therefore be taken advantage of in archives. The widely spread MP3 format suffers from losses by compression and is completely unsuitable for archiving valuable digital audio data – that said, it remains the only audio data storage method in as much as a quarter of the institutions that archive digitized sounds! In a third of digital audio archives we come across a combination of the two formats mentioned above, and rarely with other formats, such as the traditional Audio CD, MP4, OGG and WMA.

Digital video data features in the collection documentation of as few as, at the most, 17% of institutions. The discrepancy relative to the stated digital video camera ownership (29%) is probably due to the more costly process of the conversion of raw digital video data into the form of a video document. On the most part, the digital video cameras owned by the museums are relatively idle, at least as a means of digital documentation of collections. The most frequent video format is AVI, used as a single or main format by 69% of institutions working with digital video. It is followed by the MPEG format and lagging behind are other formats (codecs): DV, WMV, DVD, DV-Cam, etc. From the methodological viewpoint, the formats recommendable for archiving purposes are the two most frequently used as they have a perspective of adequate software support in the future.

Data storage media

Data storage media classification was not based on the format of the stored data as this indicator is technically irrelevant. The questionnaire set out to establish the number of the individual media which can then serve in finding the approximate volume of stored data. The basic

data storage medium in contemporary information technology devices is the hard disk although in the museums it is still approached as a tool designed exclusively for saving operational and working data. Paradoxically, probably due to the above approach, only 62 museums list hard disks among their data storage media; one municipal museum even insists they have no hard disk available! Over all, the respondent institutions admit using as few as up to two hundred hard disks, a number which is an underestimate given the above described methodological mistake in approaching the questionnaire.

Probably unjustifiably the greatest importance is attributed to optical recording (burning) on CD or DVD media. Paying no heed to their sensitivity to observing storing conditions, relatively quick ageing and a lack of efficiency in handling, the respondent institutions store more than 5 700 units of data storage media of the CD type (CD-ROM, AudioCD, CD-RW...) and approximately 800 DVD units of various types. Both optical media types therefore hold roughly 5.5 TB of data of, more or less, an archive nature. The net data volume is likely to be less as the fixed medium capacity cannot be fully used and some media are filled only partly for organization reasons.

The last category of digital media under survey was digital data tapes. Presumably, the responses revealed unawareness on the part of museum staff as to what a digital record on a magnetic tape is. As a result, the total number of more than 700 units stored in 21 institutions evidently includes, apart from DAT tapes for data back-up, a large quantity of analogue audio recordings and video cassettes. The overall volume of the stored data is impossible to quantify due to the non-uniform capacity of the different types of tapes.

Other media favoured by the museums in data storage were external hard disks (two respondents), FLASH memory modules, unspecified optical disks, ZIP media and disk arrays (or storing data on a server).

Data archives

A very important factor influencing the durability of digital data in museums (and memory institutions in general) is an adequate strategy of backing up and archiving copies of data files. In the same way that we encounter, even today, a

great disproportion in information technology equipment across the individual institutions, the conditions in the institutional provision for the safety of data are vastly different. There is some ground for believing that responses in a questionnaire aimed at the Czech Ministry of Culture may obtain information with a slightly positive bias and the described procedures are often wishful thinking and a dreamed up goal towards which it is advisable to aim. However, we do keep noticing quite vague answers such as "back-up" or "copied on diskettes and stored"! As for the large institutions, the idea of the need to back up and ensure the safety of the back-up media is fairly clear, many of those mention (and practice?) creating multiple back-up copies and keeping the back-up media at geographically separate sites – most often branches of the institution or other buildings managed by the museum's provider/administrator (the building of the regional or municipal office, library, etc.). Notes often disclose the lamentations of the staff who realize the often sad state of data archives in their institution - "...and permanent humidity approx. 65 to 75 per cent in our damp ground-floor archive" or "if we are lucky we will live to see a place for data storage before we lose some data". To sum up, data safety in Czech museums is still a far cry from the desired state and it is here that CITeM as a methodological unit which is trusted by the museums can contribute by proposing efficient solutions tailor-made to the local conditions in the different institutions.

Digital data publication

Next to the indisputable use of digital data for record keeping, safety and research/documentation purposes, it is its use for on-line publication that is constantly gaining in importance. Almost a quarter (24%) of the respondents provide a form of access to digital data describing the collections on their internal computer network (LAN), which is quite high bearing in mind that small museums often do not have the necessary network infrastructure available and some organizations are equipped with a single computer only. The volume of internally shared data is also surprising – more than 884 thousand records concerning the collection items. The publication of collection-related data on the Internet is approached by the museums with some caution. One of the

detering reasons may be the technically complicated procedure, and a frequent argument is a worry that hitherto 'unknown' valuable items come into the public domain and could, therefore, become a potential target for theft. Only nine (5%) of the respondents have by now completed the long journey which leads to the publication of at least part of their collections in the form of an Internet presentation, thus providing digital access to more than 170 thousand collection items.

The weakness of the major collection management systems is the lack of features for the easy and straightforward publication of digitized records on the web. The BACH program completely leaves it up to the user as to how he will get his data on the Internet. Demus developers offer a generic application – ProMuS - capable of generating dynamic web content from any tabular data form, including a Demus datafile. Nevertheless, even ProMuS requires skilful settings by the server administrator. As a result, collection data of respondents using exclusively BACH is very hard to find on the Internet, while in the case of exclusive or parallel users of Demus they are six museums and one gallery.

A more frequent method of publication is to issue a multi-media title on CD-ROM. This method has been tested in practise by as many as 26 institutions (15%), although most of them (69%) put out only a single title which in small museums is often a vehicle to present the institution as a whole, or is incorporated into the tourist marketing campaign of the town which also runs the museum. Two museums are currently working on their first title on CD-ROM, two museums have issued more than 10 titles. The number of copies of publications on CD varies from several dozen units to a standard of 500 units.

A total of 52 institutions (31%) invites their Internet visitors to an on-line tour of their permanent exhibitions. The simple form consists of pages with commented photographs of the exhibition rooms; a more sophisticated version

of an Internet museum offers an interactive and attractive walk through a virtual museum in a simulated cyberspace. Preparing such a presentation, and in smaller museums even ICT administration itself, are beyond the powers of the staff and a total of 23% of the respondent institutions use the services of specialized firms or are assisted by an external specialist in undertaking some jobs related, for example, to server maintenance.

As far as digital documentation is concerned, in most institutions there is no fixed assignment for the member of staff who would do the job. After 1989 some (mostly district) museums went through a "streamlining" phase and many do not even have a photographer. A standard procedure is to assign researchers with digitization work related to documenting the part of collections they are in charge of. The obvious advantage of this system is the savings in jobs, but it is virtually impossible to calculate the partial or total number of man/hours spent on digitization. Many respondents realize that and felt the need to comment on this in the notes to this questionnaire item. Based on the survey only a very rough estimate of staff members participating at least part-time of 250-300 people – mostly researchers and documentors – can be made.

Using the methodological support provided by CITeM

Most of the respondent institutions are at least partly informed on the existence and content of the work of CITeM and a great part of them directly uses the services provided. Alongside the users of Demus, which is developed in CITeM, they also include users of other documentation systems. This is in line with the mission of the methodological centre, which should be of service to all the museums in the Czech Republic.

Consultancy on ICT matters is used by 39% of institutions, various forms of training by 26% and data conversions between Demus and other collection management systems by 17%.

Source cited:

Žalman Jirí (2005): „Česká muzea" - projekt virtuální expozice českého muzejnictví. - In: Archivy, knihovny, muzea v digitálním světě, 2005. [proceedings of conference, available on-line at <http://kzv.kkvysociny.cz/archiv.aspx?id=203&idr=1&idci=1>]