The Electronic Presentation of a Scholarly Collection Catalogue: An Oberlin Case Study

Jenny Squires Wilker, Allen Memorial Art Museum, Oberlin College, USA*

Abstract

This paper presents the issues involved in presenting scholarly art historical content via a digital medium, and describes the decisions made and procedures taken to arrive at our goal of a published collection catalogue on CD-ROM. The advantages and disadvantages of student programmers and other specifics of the Oberlin project should aid other institutions in their publication choices.

Keywords

catalogue; CD-ROM; multimedia; authoring; collection; database; student programmers, V12, Director

The exceptional collection of fine art at the Allen Memorial Art Museum (AMAM) at Oberlin College is highly regarded by art historians and connoisseurs worldwide. This reputation is based not only on the quality of the works themselves, but also on the regular dissemination of scholarly information on those works, the twice-yearly AMAM *Bulletins* and the two well-known collection catalogues by Wolfgang Stechow, painting and sculpture (1967) and drawings and watercolors (1976). In 1991 the museum made the decision to build on the impressive (for their time) achievements of the Stechow catalogues and publish a new systematic catalogue in order to present newly acquired and formerly unpublished works, and to incorporate new research on those previously published. Partly due to the prohibitive costs of a full-color print publication and partly due to the intriguing possibilities presented by digital technology, the museum decided to produce the new catalogue on CD-ROM.[1] The Andrew W. Mellon Foundation generously funded this project.

Masterworks for Learning: A College Collection Catalogue, the AMAM CD-ROM, contains 170 full scholarly entries on the most important works of art in the collection. Each entry includes a 500-word discussion of the work, exhibition history, bibliography, technical data, provenance, and a larger-than-screen-size image. Selected objects and scrolls are presented three-dimensionally. To facilitate further research, each entry can be printed out in full, along with a black-and-white reproduction of the work. Label information on the entire AMAM collection of some 11,200 art works is also available through a searchable database.

Since the AMAM's primary mission is to function as a teaching museum, *Masterworks* uses the multimedia potential of the CD-ROM medium to reach out to the Oberlin student, as well as the professional art historian. Although the emphasis throughout is on scholarly content over glitzy technology and general audience entertainment, there is a presentation section in which music, videos, and voice-overs introduce the museum's history, architecture, teaching programs, and behind-the-scenes activities. In the catalogue itself, each object and its context is introduced clearly and succinctly, along with a short artist biography and definitions of technical terms. Scholarly details, such as discussions of attribution or provenance, are always available, but at a deeper level.

The teaching mission of the museum was also key in the decision to use the creation of the CD-ROM as a learning opportunity for Oberlin undergraduates. A team of computer science, electronic music, and art history majors gained professional multimedia experience in the CD-ROM project, along with an intense exposure to the AMAM collection, museum practices, and art historical issues.

This paper discusses the development of the catalogue content (scholarly research and writing, object condition surveys, and photography); this content, however, would have been largely equivalent regardless of the publication method. Discussed in greater detail will be the program specifications; advantages and disadvantages of working with student programmers; choice of authoring software; database programming and record updating; formatting and displaying the catalogue entries; producing the multimedia introduction; testing, debugging and evaluation; and comparative costs.

Development of Catalogue Content

The content of the catalogue is the real, and lasting, value of *Masterworks for Learning*. Before I arrived at the AMAM, the museum staff and art historians on the faculty had, after many passionate discussions, selected a list of approximately 325 works that were deemed absolutely essential to the catalogue. Given the time frame of the project—two years—this list was divided into two groups of entries: 170 in Phase I, to be completed by September 1, 1997, and 125 in Phase II, by January 1, 1999. (The following discussion describes the development of the Phase I publication.) All Phase I works were next examined by conservators whose reports have been integrated into the Technical Data sections of the entries. All works were photographed by Oberlin photographer John Seyfried; scanning of the resultant 4 x 5 color transparencies was done by Luna Imaging, Inc.

The bulk of the 170 entries have been researched and written by the curators at the museum, by art historians on the faculty at Oberlin, and by myself. To meet the publication deadline, however, we needed to tap the expertise of additional scholars. As Oberlin has produced many outstanding art historians over the years, we decided to seek out alumni/alumnae in selected areas, many of whom had their first exposure to art and art history in this museum. The result of our search is a rich chorus of thirty-one voices, each one speaking to a particular work or group of works on which they are extremely knowledgeable.

Several students participated, as part of their work/study financial aid package, in the development of the catalogue content: by researching the objects and the history of the museum, doing a limited amount of writing, scanning and sizing images, proofreading and formatting text, locating comparative images, securing reproduction rights, and performing administrative tasks.

Anne F. Moore, curators Marjorie E. Wieseman (Western Art before 1850), Amy Kurlander (Modern and Contemporary Art), and Charles Mason (Asian Art), and myself spent many hours discussing what the catalogue entries should consist of and how they should be displayed. We wanted them to maintain the scholarly standards set by Stechow (and other systematic catalogues, such as those of the National Gallery, London), but we also wanted to add at least some background material for non scholars. Wieseman drew up a set of guidelines for the entries that forms the basis for the scholarly content of all entries.

Program Specifications

To meet our desire to disseminate the CD-ROM catalogue to all students and faculty members on campus, the basic technical requirement was that the software had to function on a Mac Performa with 8 MB Ram.[2] Secondly, the application and data had to require less than 650 MB storage space to fit onto a CD-ROM. We allowed 325 MB per entry images (1 MB each), plus around 70 MB for small images of comparative works for the catalogue and for museum shots for the introduction. We also allowed 60 MB for one minute of video; and 60 MB for five minutes of audio. (The specifications for image sizes were maintained, but we have included considerably more video than planned and less audio.)

We wanted the entries to page like a book, but with added toolbar links to major sub-topics (provenance, literature, exhibition history, technical data, artist biography), to the complete collection database, and to other catalogue entries. We also wanted to use hypertext links to more idiosyncratic material (footnotes, glossary definitions, detail shots, and comparative images). Footnote and reference style was considered at length,[3] as was image size and resolution, but it was only after editing and formatting several entries and actualizing the software that the final entry format began to emerge.

In addition to the catalogue entry interface, we wanted a tool for searching the entire collection,[4] and a way to display short label information about all objects. We pondered the ways users might wish to search through the collection and decided that eight areas—artist/maker, nationality, title, date, object type, medium/materials, accession number, and donor—were essential, and that key word, provenance, exhibition history, and other searches might have to wait for Phase II.

We also considered many ideas for the introduction, often blurring content with presentation and new technological possibilities. We knew we wanted to focus on the museum as a teaching collection, and on how learning occurs at the museum. We also wanted to stress museum procedures and collective effort, rather than particular individuals or positions. But it was not until the students began work on the project that a structure began to emerge. This initial haziness about the introduction reflects the fact that the director and the curators were, and are, primarily concerned with the catalogue scholarship. They did not want the AMAM CD-ROM to emulate a general audience entertainment product.

Working with Student Programmers

Although we investigated other possibilities, including off-the-shelf products (there aren't any appropriate ones) and outside developers,[5] we decided that the students of Oberlin represented our best resource for software development. Not only would this approach offer students an opportunity to do some highly creative work on a real-world project, it would also give non-art majors serious exposure to the museum and its collection. Over the two-year period seven students from the departments of art, English, computer science, and the Conservatory have participated in the project. The independent thought, creativity, and maturity exhibited by these students are partly due to policies and programs at Oberlin that encourage students to take an active role in designing their own educational programs. Oberlin also offers many avenues into the new technologies besides the computer science major: the minor in Third Stream Computing (Computing in the Liberal Arts) that trains students in multimedia authoring and Web site construction; the major in Technology in Music and Related Arts (TIMARA) at the Conservatory; courses in time-based media taught in the art department; and multimedia courses in the psychology department. None of the members of the CD-ROM team were narrowly technical; several had undertaken double majors—history and computer science, music and computer science, or philosophy and computer science—and all were interested in liberal arts applications for computers. Although some of the student programmers were paid, several chose to participate in the project for credit as an independent study (private reading) and/or in a Winter Term course.

The core group of five student programmers was consolidated in a Winter Term project in January 1996. To introduce them to the museum, its issues, and points of view, these students were first given a series of tours of the museum, both in the galleries and behind the scenes, by curators, registrar, and the education office. We also invited a professional software consultant, Stephen Toney of Systems Planning, to join in the planning sessions and help the students structure the project, establish goals and a working schedule.[6]

After one or two general brainstorming sessions, the students divided naturally and easily into two groups. One group, under the supervision of John Appley of the Oberlin College Office of Communications, was most interested in developing the multimedia introduction; this group included a Conservatory major and electronic musician (Kymm Serrano), a philosophy major with a computer science minor (Boris Kizhelsteyn), and an art major, with a strong interest in virtual reality (Nathan Kelly). The other group, two computer science majors under my supervision, worked on the catalogue: one student on the database (Andrew Broadstone) and the other on the user interface (David Moxon). Additional students, both art history majors, joined the multimedia group during the spring term, focusing on the architecture section (Robert Lamb and Evan Bennett).

By April 26 this team of seven students had developed a preliminary, but functional, version of *Masterworks* that we demonstrated to the AMAM director, curators, and other museum staff. While there was much left to be done, the students had designed and implemented a complex program in less than four months, despite competing demands from other courses, and other aspects of student life. At this point, we decided to hire two of the participants to work on the project as private consultants. Over the summer and fall, David Moxon '96 and Kymm Serrano '95, focused full-time on the project and turned the working version into the final product. Working together closely, they developed their separate Director movies (Serrano on the introduction, and Moxon on the catalogue) and yet kept them coordinated in functionality and design.

These Oberlin students were inventive, ingenious, and flexible problem-solvers. On the other hand they had done little professional programming before, and *always* underestimated how long a job would take. They were committed to many other projects at the same time, and were not as immediately available as a professional software developer would have been. Though extremely up-to-date on new software, and eager for me to buy it, they had not used it in a professional environment before. These at times frustrating aspects were compensated for by the energy, imagination, and long hours the students brought to the project. In working with students, and probably all professional software consultants, it is important to stress that software is meant to be brought to completion and shipped, not endlessly perfected in the timeless universe that students tend to occupy.[7] And, again, if it were not for the Oberlin students, *Masterworks* would not exist at all, certainly not in the form that it has taken.

AUTHORING SOFTWARE

The choice of programming language and/or authoring tool was researched and debated by the Winter Term group. The multimedia group was comfortable with MacroMedia Director, whereas the computer science students wanted to construct the program entirely in C++. Although the latter choice would have allowed the greatest control and flexibility, I vetoed it, due to the short time frame, the need to have a program that future students could continue developing, and the desire to expand to a cross-platform version in the second phase of the project. We did not seriously consider any of the other authoring tools, such as HyperCard (regarded as too simplistic) or Authorware (not available on campus). Director was "hot" when we made our decision, and once such a decision is made, and considerable learning time has been invested, there is little inclination to change.

NUMBERS OF COLORS

Another decision the group considered was whether to produce the publication in 256 colors with custom palettes, so that it would be viewable on older systems, or to allow for "thousands of colors" only. We opted for the latter choice, because it offered ease of programming and better images, and also because of an aversion to programming for the past rather than the future. Thousands of colors are viewable on the target system, the Performa 636.

DATABASE PROGRAMMING AND CONTENT UPDATING

Director was suitable for authoring the introduction and the display of the catalogue entries, but we needed an external program (or XObject) to perform database operations. We initially tried FileFlex, but the searches were too slow. Broadstone programmed a custom database in Mac Perl, and though it literally produced faster and faster search results each week as he refined his program, we were worried about the future of such a personalized program. Late in the semester, we discovered V12 Database Engine, produced by New Integration Media, a cross-platform database management program specifically designed to work with Director. After trying it out, Moxon converted Broadstone's custom program to a V12, which meets our needs in the short run, although its indexing capabilities are limited. He designed the user interface, with pull-down menus for searching by object type, nationality, and chronological eras, in Director (FIGURE). Search requests are passed from the Director interface to the V12 data table; the results are returned to Director for display in a result list.

The content of the database was also a major issue. The records of the museum's collection management database (Argus by Questor Systems, Inc.) were in very good order for internal use, but the data had to be reviewed by the curators before external publication on a CD-ROM. The relevant data was printed out from the Argus database, corrections were made on hard copy, and then entered into a FileMaker Pro database, which contained converted data from Argus. Moxon designed the FileMaker Pro screen for us, as well as a script that exports the corrected data into a tab-delineated file. This latter file was then imported into the V12 database for use on the CD-ROM. At the same time printouts of the corrected data were given to the registrar who updated Argus. By publication date the three databases (Argus, FMPro, and V12) will concur in all essentials; they will all need to be maintained simultaneously with new acquisitions or additional corrections.

Although the review of all collections records was a major task, it ultimately turned out to be a benefit of the CD-ROM project. Not only has our collection data as a whole been scrutinized, it is now available to outside audiences in a manageable way. Students are already responding positively to the increased collection access.

THE ENTRY INTERFACE

One field in the V12 database is used as a link to the catalogue entries. Information in this field directs the program to open the specified folder. to size and display the enclosed image (a PICT file with a maximum of 750 pixels in either direction), and to format and display the enclosed text files. The choice of layout depends on the orientation of the image, horizontal or vertical.

The entry interface also provides an icon that triggers the zoom function for the images and runs the QTVR movies for sculpture. Each screen also provides hypertext links to comparative images, glossary definitions, and footnotes. The entry toolbar allows the user to navigate around the entry, to go to all other elements of the program, to get help, and to print out the entry (via Print-O-Matic Xobject).

Moxon was able to design an interface for the entries, as well as for the database, that is consistent with the already established public image of the museum, and at the same time incorporates contemporary design elements. While not a trained graphic designer, he nevertheless has the ability to mold pixels and their "thousands of colors" into a compelling and engaging information space, one which leads the eye to important facts, tools, or related pages.

FORMATTING THE ENTRY TEXT

The preparation of text for the entries—after research, writing, editing, copy editing, and proofreading—is time-consuming. The text is broken into 12 files, one for each of the "main" entry screens, and one each for the label information (printed at the top of each screen), artist biography, provenance, technical data, literature, and exhibition history. For the "main" screens the text is also copy-fit so that it fits within the layout limitations (width and numbers of lines). All files are converted to plain text format and formatting commands (similar to HTML) are entered throughout. Glossary definitions are entered into a glossary database, as are links to comparative images and other works in the collection. An Oberlin English major (Erika Baxter, OC '98) does the bulk of this painstaking work. After she finishes filling the folder with the prepared files, she runs the program to make sure the text fits properly and that formatting is as desired. She then prints out the entry from within the program, and it is sent back to the writer and myself for final revisions.

PRODUCING THE MULTIMEDIA INTRODUCTION

Work on the introduction progressed in an evolutionary way, with much discussion, several conflicting designs, and many false starts. Kymm Serrano, who structured the introduction as it has been ultimately realized, was often frustrated by the lack of content, which was only slowly developed over time. The architecture section was designed and redesigned, written and rewritten, several times; it represents three semesters of work for one student (Rob Lamb), one semester for another (Evan Bennett), plus the time of a faculty member (Stanley Mathews). From the start there was interest in incorporating at least some QuickTime™ VR, but I was determined to keep this as a minor feature, since at worst it might not work and at best it was a "cool" feature, not a substantial scholarly contribution. Whatever we might be able to accomplish in that area would surely be surpassed in a short time by new technology. However, Kelly and Kizelshteyn persisted, and the QTVR movies that we have incorporated have been generally well received, particularly the three-dimensional views of objects.

I was surprised by the time required to produce this short introduction to the museum. The short, grainy video bits, for example, involved the coordinated efforts of many individuals and hours of Serrano's editing time. But the response from our beta testers (see below) to this part of the program has been very positive. They enjoy and learn from the videos, the QTVR, and the information on the building, so it appears to have been worth the effort. Serrano's subtly humorous sense of timing is evident throughout all the videos, music, and audios.

One of the final software deliverables linked Moxon's and Serrano's Director movies, making it possible to go back and forth between the catalogue and the presentation at any time. Within the introduction, for example, there are links to pre-programmed searches in the collection database of the acquisitions of all past directors, as well as those given or funded by major AMAM donors.

Testing, Debugging and Evaluation

In December 1996, after the FedExing of many zip cassettes back and forth between the programmers and myself, we pressed the first test CD-ROM, and it worked in all major aspects. However, we discovered unfortunately the program requires more than 8 MB RAM to function properly. The presentation movie runs on 8 MB (except for the videos), but the catalogue does not function at all, because the hefty catalogue application, the V12 database, and the 1 MB images must all be in memory at the same time. We cannot change this situation at present, but by now the target Performa 636 with 8 MB RAM is "obsolete," and students are buying systems with at least 16 MB RAM.

In March, 1997, we shipped a beta version of the CD-ROM, fully functional but with only 28 full catalogue entries, to 50 testers in other departments on campus and elsewhere. The program performed well on a variety of systems, although there were still some bugs. The most important feedback came from people who had not seen the program before; their confusion with certain sections or functions, as well as their praise, have been invaluable in the finalization of the publication. The introduction, which was honed with such difficulty, has been received positively by almost all testers. The entries and the images have also been extremely well received. Student testers had no problem using the database and enjoyed the access to the collection they didn't have before. Other testers had difficulty finding what they were looking for, or were impatient with the slow text string searches (in title, medium, and donor fields). We also received many suggestions for new features, which will be helpful in the planning of Phase II of the project. During the summer of 1997 a new computer science student (Eric Scuccimara OC '98) was hired to perform debugging and general cleanup. Baxter (the English major) continued her indefatigable formatting efforts. Several test burns were done over the summer, leading up to a final release on August 15.

Masterworks for Learning has largely achieved its original goals: it provides a scholarly catalogue of the major works in the AMAM collection, a searchable database of the entire collection, and an introduction to the museum. In Phase II we plan to enrich the program's functionality, primarily for research purposes. Keyword or subject searches; manipulation, saving, and printing of search results; capture of comparative material from other sources; user slide shows; and above all cross-platform capability will allow art historians and curators elsewhere to maximize the potential of the information contained in *Masterworks*. We would also like to index the text in the catalogue entries for more extensive searching as well (this will necessitate another change in database platform).

Comparative Costs

Does a CD-ROM represent a cost-saving method of producing a collection catalogue?

Research and writing, conservation surveys, editing, fact-checking, and proofreading costs were similar to those of a print catalogue. The cost of text-formatting ($6,400) probably represents an equivalent cost to design and page layout, although it is not as highly skilled a task and therefore less expensive per hour. We undoubtedly spent more on photography and on comparative images from other institutions than we would have for a print catalogue, simply because the cost of printing so many color images would have been prohibitive. Our imaging costs, however, were significantly lower than equivalent color separation costs in print publication (total cost for photography, imaging, and rights, $15,000). Software development, including the purchase of required hardware, was a significant additional cost not required in a print catalogue (software: $30,165; hardware: $8,618). Production costs, however, were much lower (less than $10,000), than those of a printed catalogue (estimated between $150,000-200,000). Finally, most of our costs were lower here in Oberlin, Ohio, than they would have been in a more urban location; the use of Oberlin student labor, and their use of their own equipment, was also highly beneficial to the budget.

Our total project costs were roughly the same as the production costs alone for a printed catalogue. Additional information, images, and multimedia effects, and ease of updating are other advantages to the electronic medium. But whether this format will meet the future needs of scholars and research as well as the portable, durable book only time will tell.

REFERENCES

1 Anne F. Moore, former director at the AMAM, conceived the original idea for the project; this catalogue would not exist without her efforts.

2 During 1995-96 most incoming students were purchasing Mac Performa 636CD's with 8 MB memory and double-speed CD-ROM players. We knew that future students would have systems with at least this much capability.

3 We were aided here by Katharine Whann, Assistant Editor, Systematic Catalogue, National Gallery of Art, Washington, D.C., and by Margaret Aspenwall, Editor, Metropolitan Museum of Art, New York.

4 The AMAM collection management program, Argus by Questor Systems, Inc., is far too complex for this purpose.

5 Of great help to the project was Stephen Toney, Systems Planning, who came to Oberlin during our start-up phase. Also helpful was Ben Davis, Program Manager, Communications, The Getty Art History Information Program, who shared with us material helpful in the planning process. Both these contacts were made at ICHIM '95 at San Diego, where I underwent a crash course in multimedia development.

6 Toney set up a project plan in MacProject Pro, which has been helpful in identifying troublesome timing issues, and in keeping us more or less on schedule.

7 Toney introduced the concept of the 80-hour deliverable to me at ICHIM '95. Although 80 hours of a student's time does not represent two weeks, it is a helpful standard. The idea of the deliverable, the completed task, is key.

ACKNOWLEDGEMENTS

• This paper is dedicated to the AMAM CD-ROM software team: Kymm Serrano, OC '95, presentation programmer/designer, and David Moxon, OC '96, catalogue programmer/designer; Evan Bennett, OC '96; Andrew Broadstone, OC '97; Nathan Kelly, OC '97; Boris Kizelshteyn, OC '96; and Robert Lamb, OC '97.

Linking Text and Image Databases in GENREG: A Multi-media Museum Management System at the National Museum of Denmark

Lene Rold, The National Museum of Denmark, Denmark

Abstract

The paper describes the various problems we encountered, and the solutions we found, when linking text and images from the old GENREG system in use at the Prehistoric Department to become a multimedia database management system. A frequently occurring problem is that manual documentation systems have a very different approach to «unique identification» of artefacts/objects than the approach taken by an electronic system. It became necessary not only to reconcile manual versus electronic documentation methods, but also to consider a way of classifying images according to their suitability for different uses as well as the actual use of the image in publication, scholarly documentation and research.

Keywords

Multimedia database. Unique identification; Classification; System migration

Background

The system, called GENREG, was first developed 10 years ago as a rather simple relational database system whose main purpose was to facilitate transfer of selected textual information in the museum's collections from the manual archives to databases. The retrospective database was a distributed system whose architecture conformed to a general data model, but allowed for necessary variations among collections of artifacts from Danish prehistory, medieval times, modern times, the ethnographic collection, and some specialized smaller collections. In total the retrospective base today holds textual information on 1,000,000 artifacts. Once the text base had been established, the GENREG project was enlarged with a system to record photographs of the museum objects, resulting in a base of almost 200,000 electronic images.

Around 1990 we started developing a new GENREG system based on our experiences and growing awareness of the differences between manual and electronic documentation methods. The new GENREG system has now been in use for a couple of years in documentation of new museum accessions, and we are in the process of converting the data—both text and images—from the older system to fit into the new.

Readers interested in more of the background may consult the proceedings of ICHIM '95 (see lit. 1) where the new GENREG system is described.

Obstacle 1: Curatorial documentation methods that have varied throughout 200 years

The retrospective base reflects the manual documentation at a series of times through the collection's existence—e.g. almost 200 years of curatorial museum management is represented in the changing fashions of documentation methods.

A problem occurring very often is that manual documentation systems have a quite different approach to "unique identification" of artifacts than the approach taken by an electronic system. Also, in old collections stemming back to the last century, it is certainly evident that the approach to this crucial problem has been dealt with differently over time, so that the ways of allocating inventory numbers often become a matter of opinion held by the curator responsible for documentation, or reflect ebbs and flows in the resources spent on documentation. A commonly occurring phenomenon is the "miserly" way in which curators used to deal out inventory numbers as the following example shows.

In the register of new accessions into the prehistoric collection from 1846, the curator writes:

> Inventory no. 9221: one unspecified fibula, 3 tutuli, more than one spiral arm ring and also at least one finger ring, another arm ring of different shape, a button, a sword,

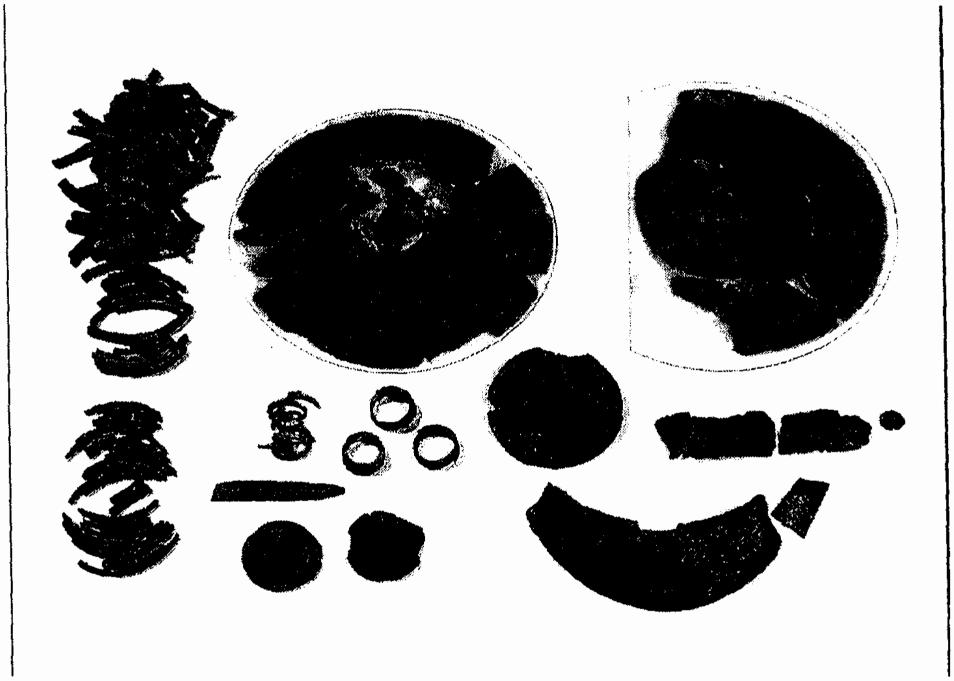

Figure 1: Some of the artifacts with the inventory number 9221. The photo shows the artifacts which presumably stem from a woman's grave. Photo: Arnold Mikkelsen.

a neck ring, a plaque worn on the belt, and a ferrule for the sword's sheath.

Inventory no. 9221 (see Figure 1) is in fact a mixture of artifacts from two graves—a man's grave and a woman's grave—from burials in a bronze age barrow, dated to 1500-1300 B.C.

Granted, the documentation is not always as messy as this example shows, but it is certainly true that inventory numbers are often used in order to keep track of the relationships among objects; for example, the relationship that the several artifacts were found in one grave.

OBSTACLE 2: A MULTIPLICITY OF ARTIFACTS UNDER ONE INVENTORY NUMBER—IMPLICIT KNOWLEDGE AND STRUCTURE: SHARED NUMBERS IMPLY SHARED FINDS, PERHAPS TIME AND SPACE SHARED.

Inventory numbers have a deeper meaning in most manual systems than identifying a specific artifact—inventory numbers may even be treated with almost religious respect, with curators reserving the very "best" numbers for outstanding artifacts. Some things become clearer when treating the inventory numbers in such ways—for one the number is easier to remember, and secondly, artifacts that share a number are easily seen to share something else, like having been together at some specific spot in time and space—something which has always been of utmost importance to archaeologists. In the prehistoric collection at least eight different numbering systems have been in use, as well as additional subnumbering systems within each of the main systems.

In relational database systems, the values of key fields are crucial for linking together information in one-to-many relations. Since a database mirrors real life, it is important to find a correspondence between the system's key values, and the real world identification of the objects documented in the base. Thus most museum management systems tend to enter inventory numbers in the basic table which is the root for all the other tables, thus equating the system's key value with the inventory number.

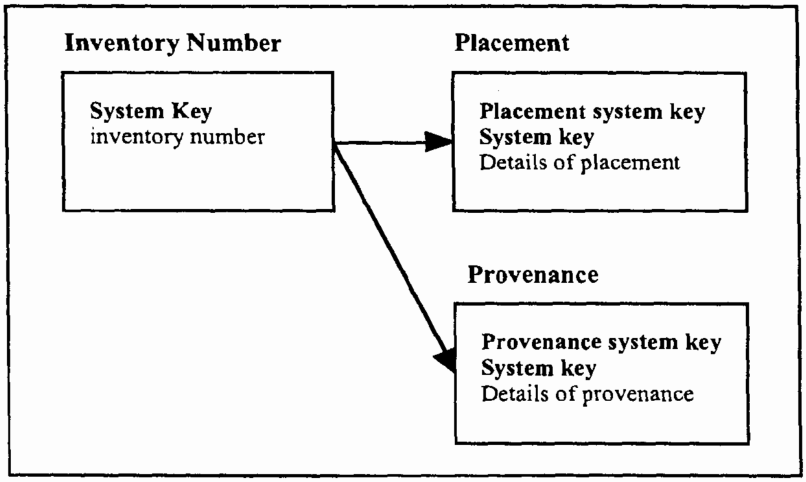

Figure 2: A simplified data model of the retrospective GENREG system in use at the Department of Prehistory, at the start of entering data into the system from manual archives.

In the retrospective GENREG database for the prehistoric collection, we chose the placement archive as the source when starting to transfer data to the base. The placement archive consists of small updated index cards with—presumably—all the artifacts represented by inventory numbers, provenance, and details on where the artifacts are placed in permanent storage or in exhibitions. The data model consisted of three tables allowing for the fact that an inventory number might include more than one artifact, and that these artifacts might be stored in different places; as well as the fact that we sometimes have more than one opinion on the provenance of artifacts (see Figure 2).

After establishing the basic database (which held about 125,000 inventory numbers), the next stage was to inspect the artifacts and enter information on—among other things—the classification of each artifact, dating, and materials; and the number of artifacts with the same classification. As the inventory number could cover anything from one unique artifact up to a container with more than five hundred pieces of flint waste, the database for this phase of the documentation project was enlarged by a new table (see Figure 3).

Figure 4 shows the result of a search in the database for inventory number, classification, and number of artifacts with a specific classification, using the inventory number 9221 as an example.

OBSTACLE 3: THE NUMBER OF ARTIFACTS IN EVEN THE BEST DATABASE RECORD CAN BE UNKNOWN

When this part of the project ended the classification table held 180,000 records, as opposed to 125,000 inventory numbers. It was not possible to determine how many artifacts the collection in reality held. Figure 5 shows one of the reasons. The artifacts are often in a very fragmented state, and in this specific example the archaeologist would only state that he believed that the fragments shown originally consisted of more that one finger ring.

The field in the database which holds information on the number of artifacts with the same inventory number and classification code is, for this reason,

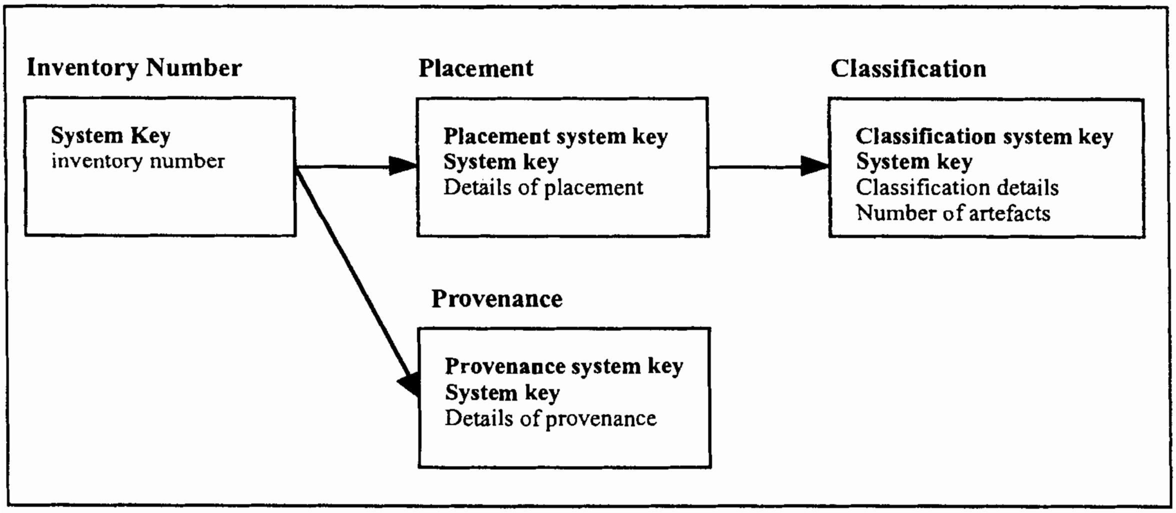

Figure 3: A simplified data model of the retrospective GENREG system at the time of classification

System Key	Inventory number	Classification	Number of artefacts
1234	9221	fibula	1
1234	9221	tutulus	3
1234	9221	spiral arm ring	>1
1234	9221	spiral finger ring	>1
1234	9221	arm ring	1
1234	9221	button	1
1234	9221	sword	1
1234	9221	neck ring	1
1234	9221	belt jewelry	2

Figure 4: Records for the same inventory number may not only consist of many types of artefacts, but also a number ofartefacts of the same type

not a number field; thus the actual number of artifacts in the collection could not be calculated by summing up the data in this field.

Obstacle 4: The textual database isn't always ready before the images are registered

At various stages of completion of the retrospective electronic documentation project, we received funds to photograph selected finds and artifacts. However, in cases when the task of entering the basic textual database had not been completed, we usually "rigged up" a photograph system which took its input from the inventory number written on the artifact (which the photographer could read), and paired this number with all the images produced of the artifact. We were quite aware that this was not a safe procedure, and expected to run into problems when the image base thus produced had to be joined to the main GENREG base.

In 1996 we received funds for an Internet publication of selected finds from the oldest part of the bronze age (see lit. 2). From the systems development point of view, this happened at a very unfortunate time, as we had by then concluded the retrospective documentation but had not yet started the conversion of data to fit into the new GENREG system. However, solving the problem of unique identification of artifacts became crucial, so during a few short weeks we had to analyze and solve the basic problems of linking text and images in the documentation system, and as well to take into account electronic publication based on this same information.

Obstacle 5: Unique identification is even harder when images are involved.

By joining the tables which held, respectively, the inventory numbers and the classifications of the artifacts, we created a new table in order to add serial numbers to the inventory numbers which held more than one artifact (see Figure 6).

The new table enabled us to identify artifacts with the same classification in the problematic finds. Most of the artifacts could, by this operation, be uniquely identified within the system by concatenating the systems key and the serial number. In the real world an artifact could be identified by inventory number and the classification term. However, there still existed a problem of identification when an inventory number held more than one artifact with the same classification. It would have been tempting to create records based on the actual number of artifacts with the same classification term. By doing this, we would be able to identify the three objects shown in Figure 7 by unique records in the base. However, since the column holding the information on number of artifacts is not a number field, we decided to wait while looking into this particular problem.

In the new table in the database system, these three tutuli are thus represented by only one record. This means that from the systems point of view there is only one object. So when we link the image part of the system directly to the newly established object table (see Figure 6), any images of these tutuli would be accompanied by a text stating that the image shows three tutuli—also images which show only one of them(!)—a problem which will be dealt with later in this paper.

Why—you may ask—do we not give each of these tutuli a new inventory number, or an addition to the original inventory number? Well, for one, you have to paint the additional number or letter on the artifact itself (to avoid losing track of it); but you may also have a problem in tracing the artifact back within the old manual archives and literature. In the old records the artifacts may be identified differently, which will invariably lead to more confusion than clarity. Finally, the work task involved in creating a new additional inventory numbering system, and carrying out the repainting of artifacts, is huge since more than 50,000 artifacts in the prehistoric department alone may pose this special problem.

Our efforts uniquely to identify all artifacts show that this is not possible, either because the artifacts may be fragmented (see Figure 5), or because it is undesirable, for instance in finds consisting of hundreds of potsherds or flint waste. In case it is of utmost interest to identify an artifact, we shall do it while proceeding with the photography project, or other manipulations of the collection. In the example of the three tutuli (see Figure 7), we could thus create a record for an individual tutulus in the text base in

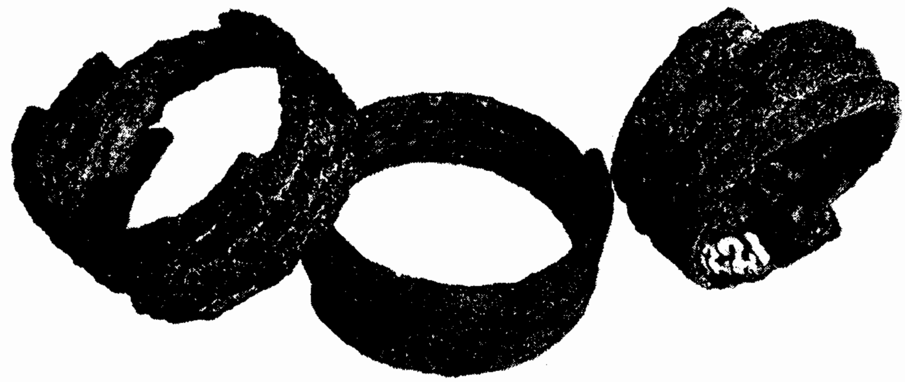

Figure 5: This images shows fragments of finger rings from inventory number 9221. The text base information linked to the images would be "more than one spiral finger ring". Photo: Arnold Mikkelsen.

case we need an image of only this one tutulus. We may also need to create a new record if we want to loan or exhibit only one of the tutuli. The problem of unique identification is not only a problem when photographing the artifacts; it is indeed crucial for many other manipulations in museum management.

OBSTACLE 6: AN IMAGE CAN BE A NECESSARY PART OF AN ARTIFACT'S IDENTIFICATION.

For the time being we can thus conclude that identification of an artifact within the system involves the old systems key and a serial number. In the real world, identification involves an inventory number, a classification and—when more than one artifact with the same classification is involved—an image. It is indeed interesting to note that images this way become an important part of the identification problem.

The reason for this is that the curator is faced with finds where the artifacts share the same inventory number written on the artifacts themselves. If this find, for instance, is inventory number 9221 he may safely be able to identify most of the artifacts according to their classification, but if the three tutuli were split up in three records then all the textual data for each of these artifacts would be identical and so only images of the individual tutulus would enable him to identify each of them.

The situation may change as the textual base in the future becomes enhanced with descriptions of the individual artifacts, such as measurements, motives, and inscriptions on the artifact. However the inventory number, the classification term, and the images were in the database first, and so have to be taken into account before any additional documentation is added. In the case that the image is a part of the identification, you need access to the image before you start adding further information into the base. This may be disturbing for museum managers used to manual systems, but we consider it of utmost importance and relevance to acknowledge that images may play such an

System Key	Serial Number	Inventory number	Classification	Number of artefacts
1234	1	9221	fibula	1
1234	2	9221	tutulus	3
1234	3	9221	spiral arm ring	>1
1234	4	9221	spiral finger ring	>1
1234	5	9221	arm ring	1
1234	6	9221	button	1
1234	7	9221	sword	1
1234	8	9221	neck ring	1
1234	9	9221	belt jewelry	2

Figure 6: By joining two tables inthe original database we created a new table. The original systems key was no longer unique, but had to be concatenated with a serial number to uniquely identify each record.

active role in future multimedia documentation systems.

OBSTACLE 7: ARTIFACTS IMAGED MANY AT A TIME, ONE AT A TIME, AND ONE PART OF AN ARTIFACT AT A TIME.

The easier access given by electronic imaging to produce images at relatively low cost will necessarily lead in practice to producing a greater number of images. Any handling of the fragile artifacts will invariably lead to further deterioration so we became aware that the imaging project, even though in it self a threat to the artifacts, was a unique opportunity of documenting our prehistoric collection. Consequently, a great many photographs were taken during the electronic imaging sessions in order not only to produce images for the Internet publication, but also images which might document the artifacts so well that much future handling might be avoided by giving access to good instructive images.

Not only did we want to have images of the "family photo" type, where a number of artifacts from the same find were photographed together (see Figure1), we also wanted "individual portraits" of artifacts (see Figure 8), and close-up images (see Figure 9).

Once an artifact had been placed in position, the photographer would experiment with light and thus produce more than one photo, where different parts of the artifact would sometimes be lighted differently in order to show certain characteristics. Further, most physical artifacts have more than one side, and the back may be just as interesting as the front. The result of all this documentation activity actually resulted in some images which were unintelligible to all but the trained archaeologist's eye, and which were utterly uninteresting if to be used for an Internet production aimed at a broader public

In order to meet the dual demands of both thorough documentation and of images which would appeal to a broader public, we needed to decide on a

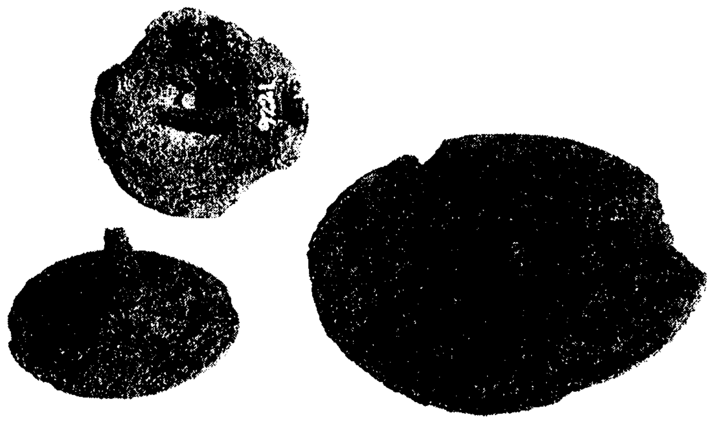

Figure 7: This images shows three tutuli from inventory number 9221. The text base information linked to the images would be "Three tutuli". Since there is only one record in the text base for these three artifacts it is impossible to derive a text for an image showing only one of the tutuli. Photo: Arnold Mikkelsen.

procedure to mark the images according to their usefulness for publication—this issue will be discussed later; we also needed to decide on a way to describe close-ups according to the "part-of" problems well known in all documentation systems.

For the time being we have resolved the problem by dividing the images into three groups:

1) Family photos. Images of finds with more than one artifact, where each object (not necessarily each artifact) is represented by a unique key in the textual database:

   a: A family consisting of only one member

   b: All individuals from the same family

   c: A subset of individuals from the same family.

2) Individual portraits, e.g. images of an individual object as represented by its unique key in the textual database. (The image of the three tutuli, see Figure 7 is thus considered an individual portrait.)

3) Details of individual objects.

The system can take care of group 1 and 2 so that the photographer does not manually have to type in information for these image groups. All the photographer has to do is to type the inventory number found on the object he is about to photograph, and the system will automatically link the unique keys of the textual database to the image table, e.g. the unique image key is linked to one or more object keys. He must, though, take into account that the information given on the screen is important in order to understand that, for instance, the three tutuli (see fig 7) are considered to be one object by the text base. This means that if he wants to photograph just one of these tutuli, the resulting image would belong to image group 3—a detail of an object—which is information that he has to type in.

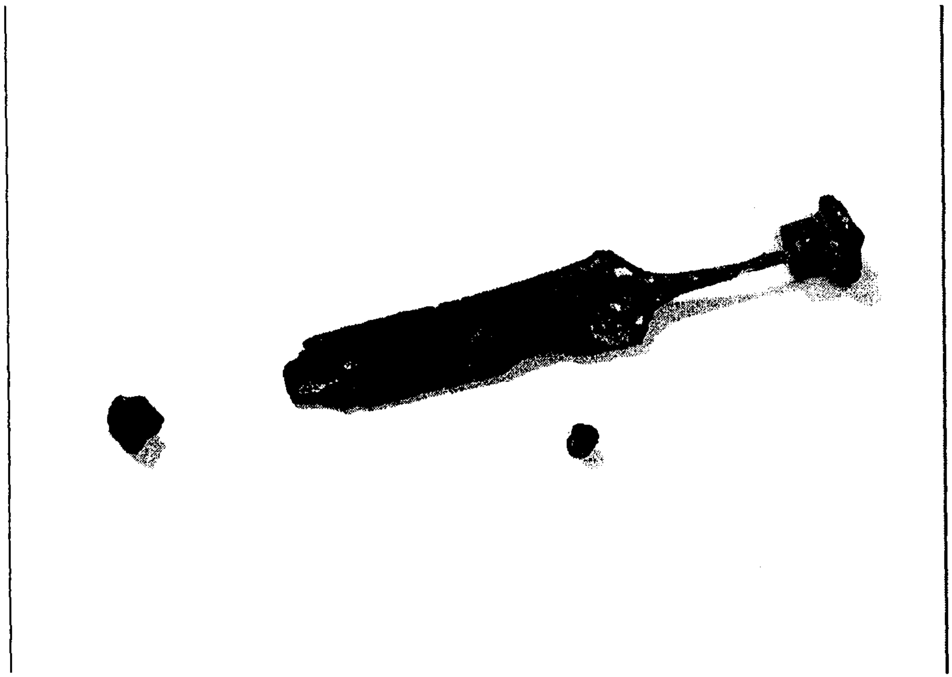

Figure 8: This image of a sword is classified as an "individual portrait" of one of the "family members" belonging to inventory number 9221. Photo: Arnold Mikkelsen.

The other solution—to make it possible for the photographer to create a new record in the text base for one of these tutuli—has not been implemented for this specific project. The reason is that we lacked resources, and in any case we were about to transfer the data to the new GENREG system, where this possibility has already been implemented for future work on enriching the collection's text and image documentation.

The third group of images—details of individual objects—is an interesting group closely connected to the "part-of" problems of documentation of objects. Assuming we actually had information in the textual database on motives, descriptions, and other "part-of" information on individual objects, we would most certainly want to link images of the group 3 type directly to this type of textual information. Some of our retrospective databases have information of this kind in the textual part of the base, and the new GENREG system deals effectively with the "part-of" problems. We have therefore discussed whether, in the new GENREG system, we should continue linking images to the identification of artifacts, or instead make a linkage between the deeper embedded information on "part-of" descriptions.

This decision is not an easy one. We certainly have no problem in programming such a solution, but we have to be aware that as the system becomes more sophisticated in dealing with various documentation problems, it also demands more specialist knowledge on the user's part. It is not evident that a photographer can distinguish between two different motives on an altarpiece. It may indeed be very difficult even for a specialist to distinguish between depictions of different saints—just to give an example. So we are actually faced with a problem which cannot be solved by curatorial and documentary logic, but has to be considered from an organizational point of view—e.g. who are actually in charge of the documentation, and are these people the same who actually enter the information in the electronic system.

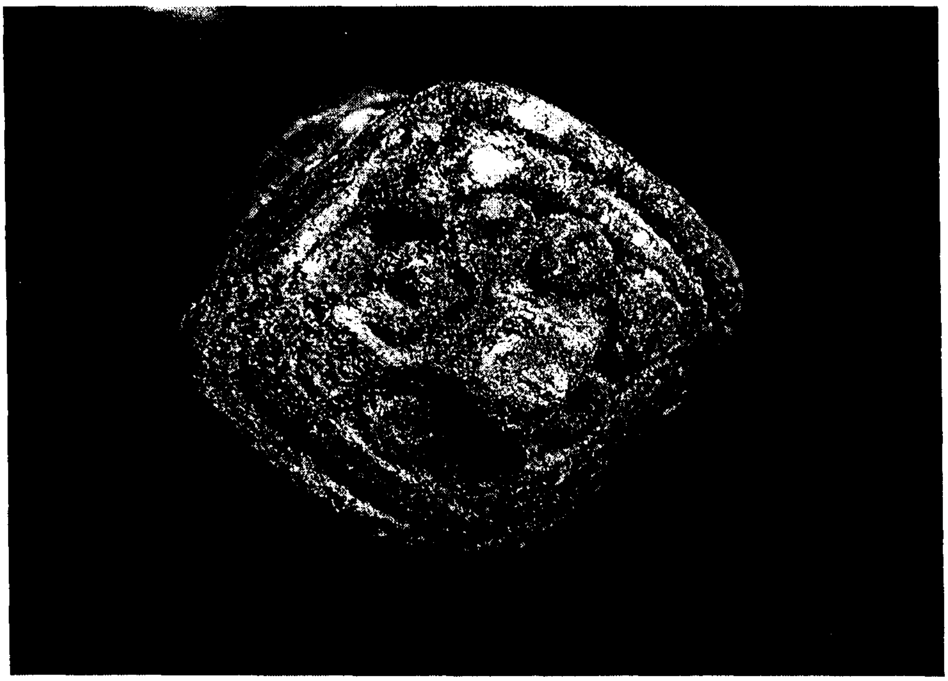

Figure 9: The knob of the sword shown in Figure 8. The image is classified as a "detail of an individual". Photo: Arnold Mikkelsen.

OBSTACLE 8: IMAGES HAVE DIVERSE USES INCLUDING EXHIBITION, PUBLICATION, THE INTERNET AND SCHOLARLY DOCUMENTATION. It is important to realize that you expect different performances from different documentation methods—particularly since one of the often-mentioned advantages of electronic systems is that you have access to and can handle huge amounts of information. You expect a text and image base to be a bank of information from which you may electronically select and manipulate data for electronic publications, virtual museum exhibitions, for research, for in-loan/out-loan management, and for the museum home page—just to mention a few uses that come readily into mind.

In fact you may select and manipulate data from multimedia databases, but more often than not it turns out that selection becomes a tedious manual process, rather than executing an elegant SQL-statement. For one, human selection may be necessary if the database holds many images of each artifact, as not all of these images may be suitable for publication. Such selection cannot safely be left for the system to decide upon at random, but needs to be done manually unless you can define a way to classify the images according to their suitability for the uses you have in mind.

We were aware of this fact at the time of photographing the artifacts for the Internet publication mentioned above, so we stored information on the intended use with each image, as well as classifying the images into "family photos", "individual portraits" and "details of individuals".

In the future this information could be useful because we know that the selected images have what you might call 'public appeal,' so we can let the machine point the images out for re-use in future similar publications. We intend to store information on all actual uses of the images in order to develop a historic record on usage information. At some point we shall probably decide that showing the same images again and again in different publications becomes tedious to the audience, and therefore set about selecting some of the other available images—a process, though, which at the time being can not be done by machine power alone.

REFERENCES

Lit.1: Lene Rold: GENREG. A simple and flexible system for object registration at the National Museum of Denmark. *Proceedings of ICHIM '95*. 1995.

Lit. 2: National Museum of Denmark: Gods & Graves. *www.natmus.min.dk/kulturnet/index.htm*. 1997.

