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,

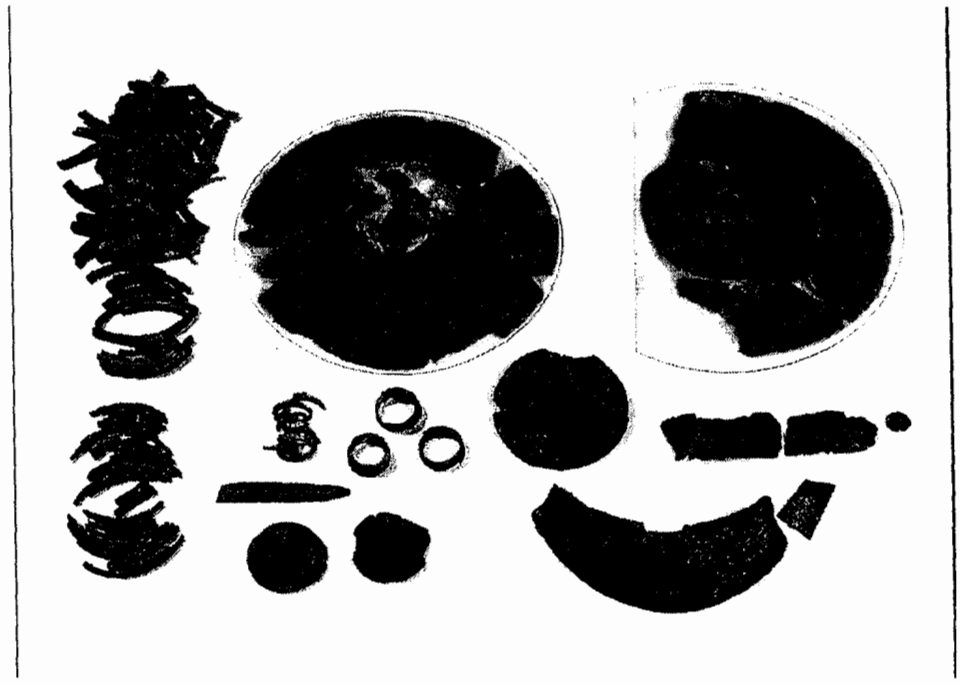

Figure 1: Some of the artifacts with the inventory number 9221. The photo shows the artifacts which presumably stem from a woman's grave. Photo: Arnold Mikkelsen.

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.

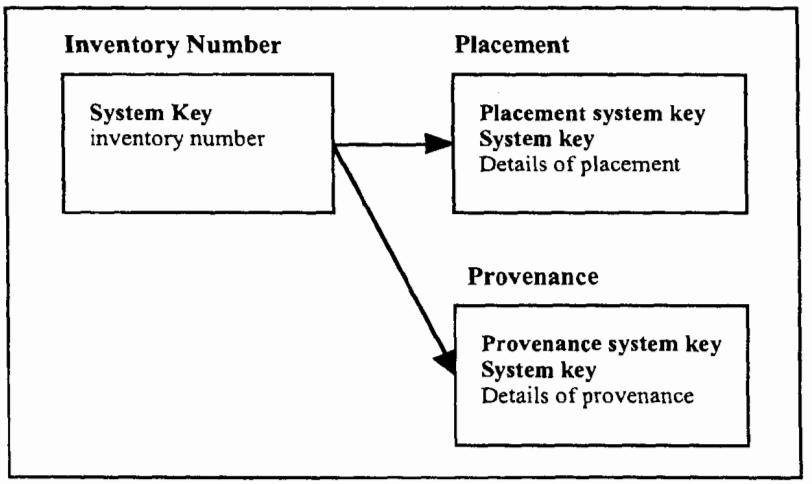

Figure 2: A simplified data model of the retrospective GENREG system in use at the Department of Prehistory, at the start of entering data into the system from manual archives.

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,

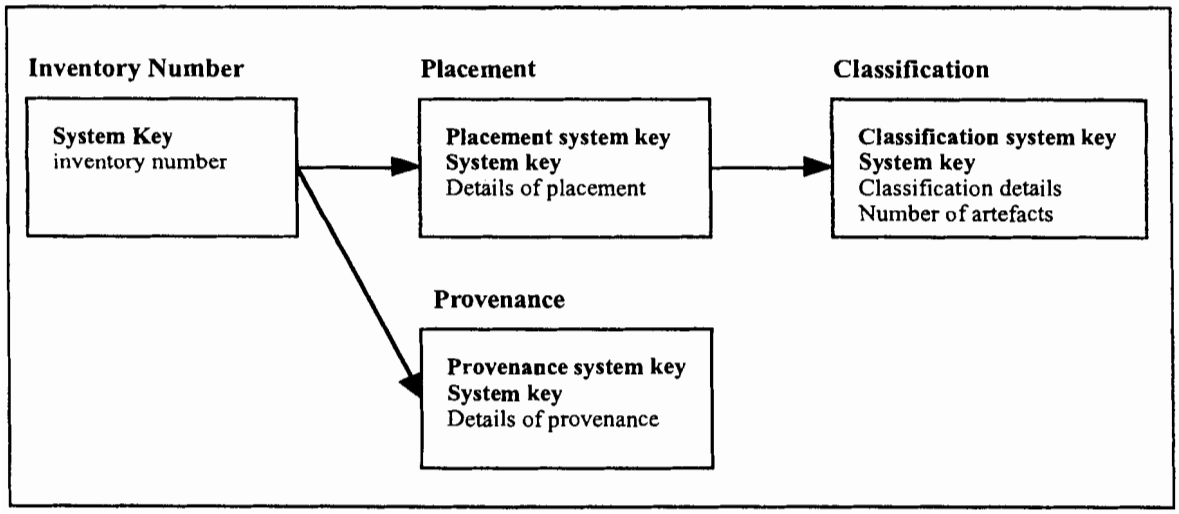

Figure 3: A simplified data model of the retrospective GENREG system at the time of classification

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

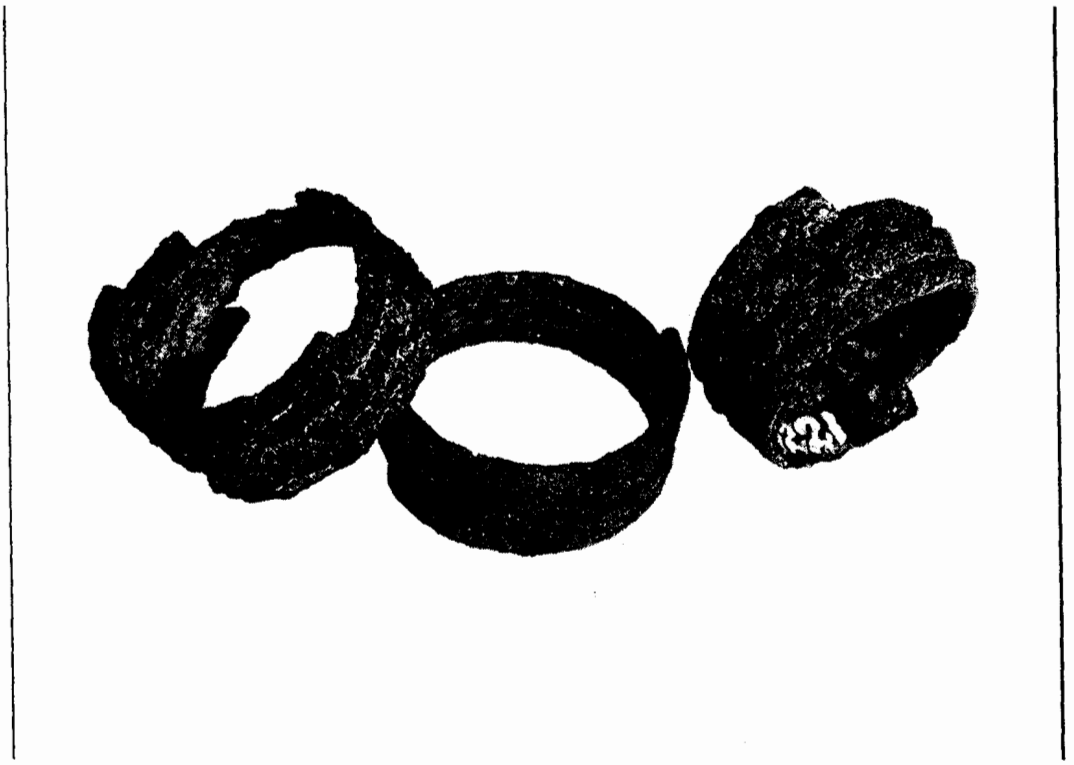

Figure 5: This images shows fragments of finger rings from inventory number 9221. The text base information linked to the images would be "more than one spiral finger ring". Photo: Arnold Mikkelsen.

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

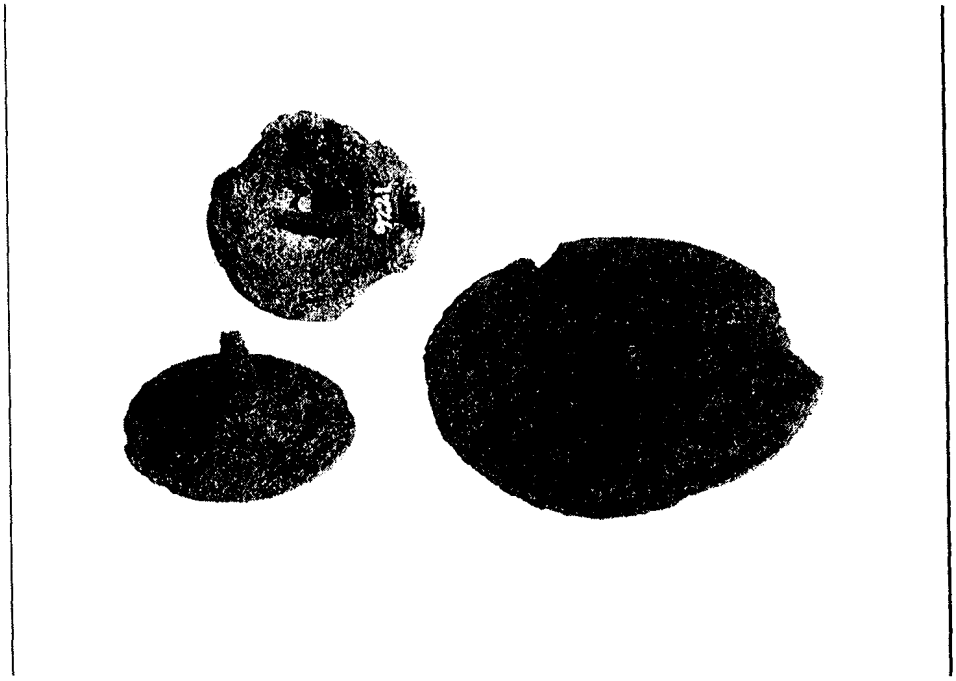

Figure 7: This images shows three tutuli from inventory number 9221. The text base information linked to the images would be "Three tutuli". Since there is only one record in the text base for these three artifacts it is impossible to derive a text for an image showing only one of the tutuli. Photo: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

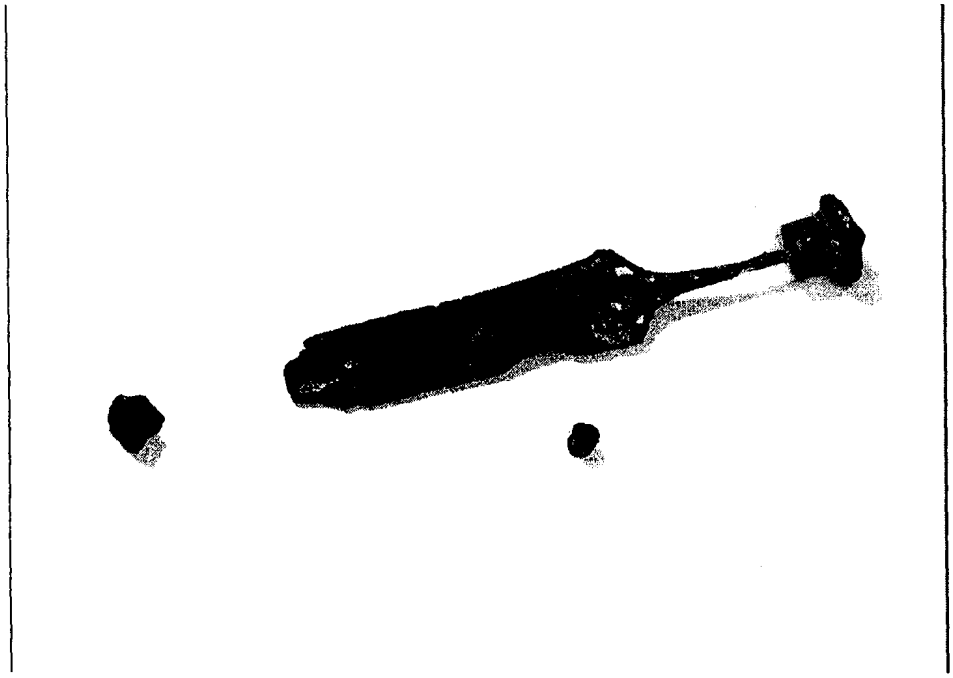

Figure 8: This image of a sword is classified as an "individual portrait" of one of the "family members" belonging to inventory number 9221. Photo: Arnold Mikkelsen.

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.

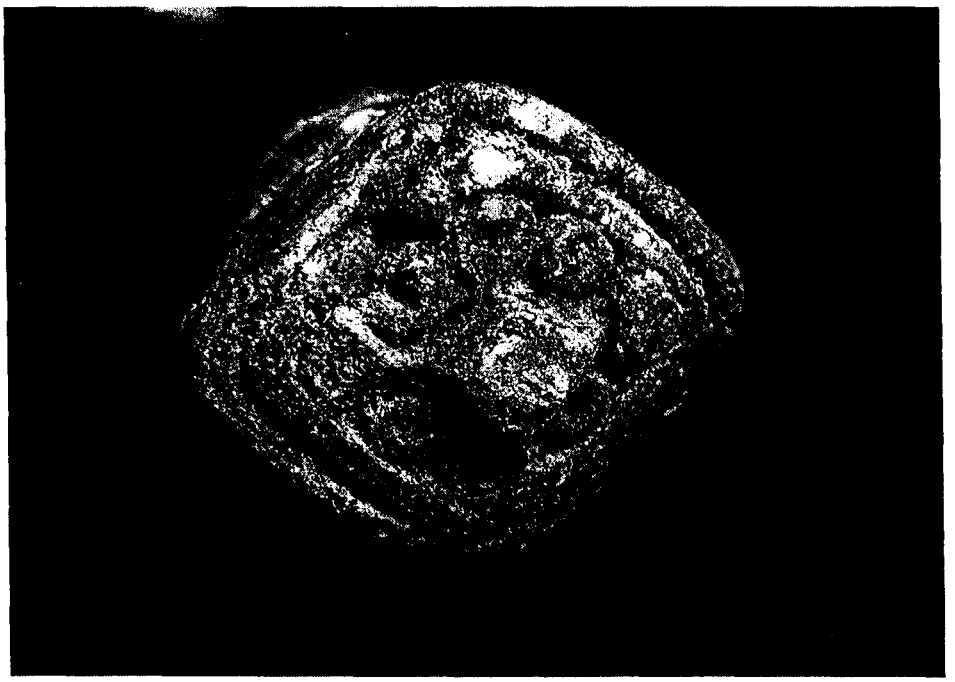

Figure 9: The knob of the sword shown in Figure 8. The image is classified as a "detail of an individual". Photo: Arnold Mikkelsen.

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.