National extensive databases in Norway

– pitfalls in a bright future

Espen Uleberg and Mieko Matsumoto

Museum of Cultural History, University of Oslo, Norway

Abstract

Extensive databases are becoming more available to researchers, students and the general public. A national database makes it more possible to have an overview of the totality. Information is easier accessible, but the amounts and quality of the information creates new challenges. The system of university education in Norway has changed. The new system together with the new availability of material may change education and research style for both experienced and amateur archaeologists. In this paper, we concentrate on the consequences for research and education but also have a quick glance at cultural heritage management and public access.

***Key words**: Norway, databases, research, education, cultural heritage, public access*

1 BACKGROUND

There are five archaeological museums in Norway (see fig. 1). All of them are university museums, and together they can be seen as a distributed national museum. The first move towards a common database system for these museums was made when the museums conjointly appointed a committee which in 1988 presented a schema for archaeological databases in Norway. It was used for the sites and monuments records, and the tagging schemata and databases in use for the museum collections today are also built on this proposal[1].

The next development was the start of the Documentation project (*DokPro*) (www.dokpro.uio.no) in 1992. It was as a

Figure 1. The five archaeological university museums in Norway

[1] Einar Østmo, "Arkeologiens arkiver på data", in *Fra skuff til skjerm. Om universitetenes databaser for språk og kultur*, ed. Knut Aukrust and Bjarne Hodne (Oslo: Universitetsforlaget, 1998), 58–79

cooperation between the historical-philosophical faculties at the then four universities in Norway. One had realized that it was necessary to make the vast amounts of information stored at institutes and university museums more accessible to researchers and the general public[2]. The extensive material was as diverse as archaeological and numismatic collections, place name registers, Norwegian mediaeval documents and folk music. The main strategy was to digitize. At this time, the internet was still new, and hardly anyone could foresee the possibilities that would open up with this technology. *DokPro* was a pioneer in this field, and was the first to use SGML formatting on a large corpus of archaeological text. The SGML-markup was presented as early as 1996 at the CAA conference in Iasi in Romania [3].

From 1998, the university museums continued the cooperation in the Museum project (*MusPro*) (www.muspro.uio.no). *MusPro* included also the university museums' natural history collections. The projects were replaced by the permanent organization *MUSIT* (MUSeum IT) (www.musit.uio.no) in 2007. *MUSIT* develops and maintains the databases for the Norwegian university museums. The archaeological museum in Stavanger joins *MUSIT* from 2009.

The four university museums with archaeological collections concentrated on digitizing different parts of their archives, but all of them converted the artefact catalogues. In addition, photos and topographical archives are in varying degree digitalized. The artefact catalogue for the Museum of Cultural History, the University of Oslo (*KHM*) has been on the internet since 1996. In the autumn of 2008 the *MUSIT* version of the web portal was opened (www.unimus.no). More than 120 000 photos from the museums can be seen here. At the beginning of 2009, the portal was extended with the artefact catalogues of the four museums. It now refers to more than two million artefacts, though the level of detail varies greatly.

Norwegian archaeology has a tradition of annually publishing the new museum acquisitions [4]. This started in the yearbook of 1866 for the Society for the Preservation of Norwegian Ancient Monuments. Here, the catalogues of interesting archaeological finds at all museums from that year were presented[5]. At *KHM*, the last printed version of the catalogue is the finds registered in 1999. From then on, the complete catalogues will be published through the databases, continuing the tradition of presenting the new museum acquisitions for researchers and the interested public.

Over the years, many of the published catalogues were not easily available anymore, and from the start of *DokPro*, a main goal was to make the material digitally available for researchers, students and the general public. The project started with converting the printed catalogues and added handwritten texts from the main catalogue. The electronic version has in this way become the first complete publication of the artefact catalogues. In a time when almost nothing could be found digitally, the conversion itself was important. Electronic publishing would remove the need of physically visiting the archives, and in that way contribute to democratization of knowledge[6].

[2] Bjarne Hodne "Dokumentasjonsprosjektet: historikk, målsetting og utfordringer", in *Fra skuff til skjerm. Om universitetenes databaser for språk og kultur*, ed. Knut Aukrust and Bjarne Hodne (Oslo: Universitetsforlaget, 1998), 13–18

[3] Holmen, Jon & Espen Uleberg. "Getting the most out of it - SGML-encoding of archaeological texts". Paper given at CAA'96 Iasi, Romania, 1996. www.dokpro.uio.no/engelsk/text/getting_most_out_of_it.html (accessed May 25, 2009)

[4] Bjørn Hougen, "Tilvekstfortegnelsene. *Arkeologiens Diplomatarium Norvegicum*." Universitetets Oldsaksamling. Årbok 1945–1948, (Oslo: Universitetets Oldsaksamling, 1950), VII–XII.

[5] Einar Østmo, "Register over trykte tilvekster av norske oldsaker". *Universitetets Oldsaksamling. Årbok 1997/1998*, (Oslo: Universitetets Oldsaksamling), 1999, 177–190

[6] Christian Emil Smith Ore " Hvordan lage databaser for språk- og kulturfag", in *Fra skuff til skjerm. Om universitetenes databaser for språk og kultur*, ed. Knut Aukrust and Bjarne Hodne (Oslo: Universitetsforlaget), 13–18, 1998

Even though the material was made available for the general public, the focus has been on databases for research. One aspect of this is that normalization has not been a priority. This means that two objects of the same type can be named by several different terms. For instance would a flint sickle be listed as a flint saw when described in the nineteenth century. Because of this, one needs more knowledge to be able to query the databases. The original artefact catalogues (see fig. 2) as they were written from 1829 give an insight in how the texts have varied over time and between the museums. They clearly reflect the research history as well as the interests of the individual researchers. Normalization is a necessary step to make the collections more accessible. Some terms can readily be replaced by a modern one, but in other cases it will be necessary to reclassify the objects.

It is a great achievement that the collections can be accessed on the internet. However, although the original catalogue texts are a necessary and welcome basis, they should be presented in different ways to reach wider audiences. Researchers prefer the primary sources, but to catch the interest of other users, the artefacts could be presented as part of a narrative.

2 RESEARCH

Archaeology has moved from positivism in the 1960ies to realizing that all data are created. Data does not exist of its own accord. Can it be possible, that the new technology will take us away from a deductive way of reasoning over to a more inductive one? That instead of setting up a hypothesis and try to find information that can weaken the hypotheses, one will collect vast amounts of information and induce a hypothesis from that? We do not think so. Instead, it will be possible to move faster in the spiral stairs of hermeneutics. A new idea can be corroborated or weakened through searches in the database. A new ordering of the material in a list or on a map can give new insight and turn the research in a different direction. A possible pitfall is the distance to the primary sources.

The researcher wants to be as close to the primary material as possible. It is important to know the material well, and preferably describe each piece yourself. To be able to work large amounts of artefacts, regroup them and create new understanding on the basis of this is an essential part of being an archaeologist. Today, the first encounter with archaeological material is mostly

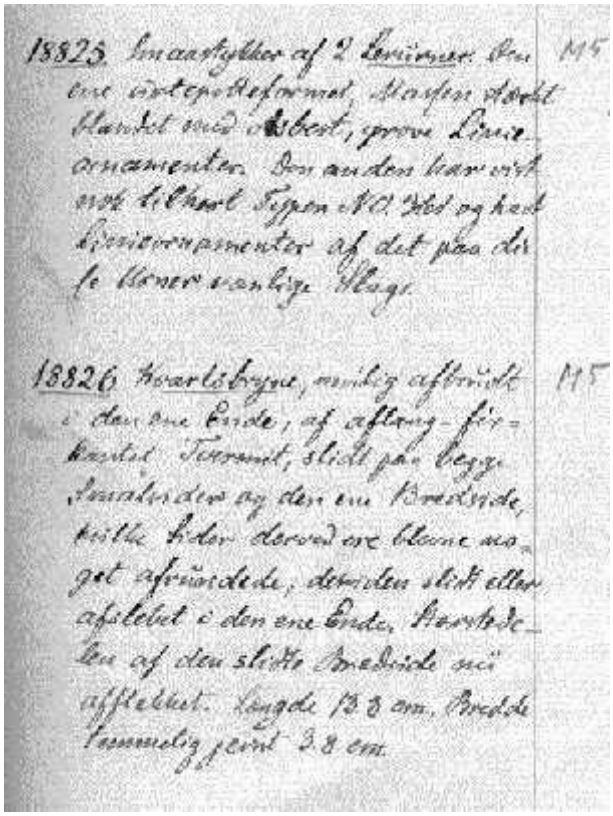

Figure 2. Catalogue text describing two of the finds from 1897. The Museum of Cultural History, University of Oslo

through a computer interface. The databases will act as a primary source for researchers and students. In this way all material are more accessible. At the same time, the distance to the real primary sources may increase. It can be possible to see a virtual presentation but not to handle the objects. Growing concern about security and wear and tear combined with better virtual presentations can be arguments to limit the access to the original collections.

In Norway, the students of archaeology now need five years to finish a master degree instead of the earlier seven to eight years education leading to a *Magister Artium* degree. This encourages more use

of secondary sources and reduced time on the study of the actual artefacts. On the other hand, a student had to spend a lot of time with artefact catalogues and in archives just to choose the artefacts they would study. In 1998, the hope was expressed that easier access to the catalogues would make artefact studies more attractive for the students[7]. Now, a first selection can be done faster. From the text and images it is possible to make a selection of material that should be looked at, and there should be more time to actually study the artefacts. Several factors are at play here – the way the institute presents the study plan for a master thesis, the working conditions in the museum's study rooms and so on.

There are several famous overviews of the prehistory for a nation or a region. They were written by highly esteemed researchers, and are regarded as authoritative presentations. Now, the sheer amount of data has become a challenge. In recent overviews, important aspects that have not been covered can always be found. Back in the 1960ies, archaeologists had defined some parts of Norwegian prehistory as almost unknown. For instance, very few Iron Age houses in Eastern Norway had been excavated up until the 1980ies. Today we know a lot about prehistoric houses from this area because the excavation methods have changed and large construction works, railroads and roads, have paid for large-scale archaeological rescue excavations. There is so much information to be found, that no one can know the totality. Hein Bjerck describes this as a shift from a collective ignorance to an individual one[8]. Earlier, the known and unknown fields were shared by all archaeologists. But perhaps this is an illusion. Perhaps the overviews were possible because choices were made as to what parts of the prehistory should be presented, and the author also had the authority to present that overview as the correct one.

Distribution maps have always been a good way to present archaeological material. A lot of effort has been required to present a distribution map as complete and precise as possible. With the advent of a national museum database with geographically referenced finds, this will change. The finds can be presented as text and as a dot on the map as soon as the catalogue is ready in the museum. They will be available, not only to researchers at the museum, but also to the general public. Anyone can choose a set of parameters, artefact type, material etc. and see the distribution on the screen. It will be possible to download data sets and do GIS analysis. The metadata should of course describe the accuracy of the position. This depends on how the artefact was found and included in the museum collection. Artefacts from well documented excavations can have very precise coordinates, while stray finds from a mountain hike will have less precise positioning. A broad distribution pattern is sufficient for some analysis, but in other cases proximity to water or other factors may demand a much more precise location.

Digital recording and easily accessible information in databases change the way we study and present the prehistory. It is also important to raise the question how digital tools influence the content of the narratives we present of the past, and whether source-criticism is lost on the way[9]. A neat presentation automatically gets a higher truth value than pieces of information in bad handwriting on a piece of paper. Digital presentations tend to promote certainty and leave out the fuzzy parts of the interpretation. But complicated sections of a structure used to be presented with neat drawings on paper earlier as well. Messy surfaces have been redrawn to support the interpretation of a site. A pitfall along the way may be that the digital presentation becomes an alluring image, covering up the interpretations and insecurity the results are based on. However, digital recording also opens up the possibility to give more people access to the

[7] Einar Østmo,. "Arkeologiens arkiver på data", in *Fra skuff til skjerm. Om universitetenes databaser for språk og kultur*, ed. Knut Aukrust and Bjarne Hodne (Oslo: Universitetsforlaget, 1998), 58–79

[8] Hein Bjartmann Bjerck, "On holistic knowledge – reflections on the occasion of Anders Hagen's 80th birthday". *Primitive Tider* (2002):164

[9] Brit Solli, "From test via sign and text to things? Or: Pimp my site" *Primitive Tider* 10 (2007/08):162.

original information with all the post holes that did not fit in and therefore left out in the final interpretation. We think it is far better with digital information with metadata that gives information about the data quality and a chance to look at the original documentation.

3 CULTURAL HERITAGE MANAGEMENT

The MUSIT database alongside the national sites and monuments database (*Askeladden*) are invaluable tools for the cultural heritage work[10]. It can be tempting to be content with the database and online maps and skip the thorough literature searches and study of the material itself. It may seem that the possibility to find facts quickly reduces the time spent for collecting information and controlling basic facts. The trust in digital sources both in terms of availability and completeness has increased enormously. At the same time, ongoing processes to decentralize cultural heritage management in Norway will leave more decisions to people with more general and less specialized background. A person familiar with a certain geographical area will be able to evaluate information related to her district. A person without general knowledge that makes it possible to evaluate the information is totally dependent on what is found on the screen. They have high expectations and become very disappointed when something turns out to be wrong.

A pitfall along the way in this case, is the user's expectancy of completeness. Users tend to expect completeness to an even larger extent than if they were using paper based archives. There is an understanding that a file might be misplaced in a paper archive. The digital archive, on the other hand, is supposed by many to be complete. One faulty query, and they will revert to the paper version. The point is that it is necessary to be critical to the sources, no matter what format they are in. No single source contains all necessary information, and no source is without fault. The digital sources can give a quick overview of sites and finds of an area. This information must be augmented with knowledge of how the area has been surveyed, historical background and other factors.

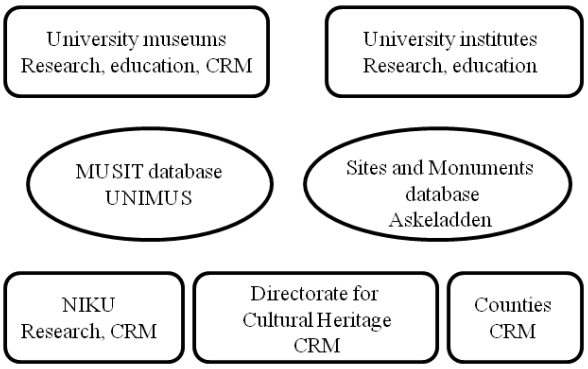

Figure 3. Databases and institutions in Norwegian archaeological research, education and cultural heritage management. (NIKU: The Norwegian Institute for Cultural Heritage Research)

4 PUBLIC ACCESS

Museum collections are presented to the general public in various forms. Archaeologists have always seen the importance of presenting the prehistory to the general public. Museum exhibitions and publications can only present parts of the collections, but the national database make the original artefact catalogues accessible for litterally everyone. This is in concurrence with the idea of democratization that is opening up public archives and making decision processes in the society more transparent. Information and knowledge shall not exist for only a small group of specialists. Accordingly, as much archaeological documentation as possible should be freely available for all. One pitfall along the way might be to be content with the presentation of the overwhelming original material.

[10] Evy Berg, "Using a GIS-based Database as a Platform for Cultural Heritage Management of Sites and Monuments in Norway", in *Digital Discovery. Exploring New Frontiers in Human Heritage. CAA2006. Computer Applications and Quantitative Methods in Archaeology*, ed Jeffrey T. Clark and Emily M. Hagemeister, (Budapest: Archaeolingua, 2007), 321–327.

Some million objects can now be found on the web pages. The artefact catalogues fulfill their original purpose as texts for archaeologist, but perhaps not so many others will appreciate them. Texts and images are a good starting point for a presentation for wider audiences, but they need to be adjusted to different user groups. *Open Archaeology* (www.intermedia.uio.no/display/arkeologi/Hjem), (*Åpen Arkeologi*) has addressed the question of how to make the Norwegian archaeological museum collections available for different audiences[11]. It has developed a concept for presenting archaeological excavations to a wider audience. Local schools have been invited to visit excavations and to use the web pages in advance and after the visit to get additional material about the local prehistory, and it has been very welcomed. *Open Archaeology* focuses on the process of excavating, data collection and give a different inroad to the artefact database at the museum – moving from objects found at the excavations to similar objects in the collection or to objects found in the same region.

The *unimus* web portal for databases developed and maintained by *MUSIT* were launched in 2008. The first web pages for the collections presents artefacts and photos with information from the original catalogues. This was a great achievement in itself, since it opened up the collections for a large number of people. This was a first step towards a web page where more information about the prehistory will be presented, illustrated with and linked to the content in the databases.

[11] Ingvild S. Andreassen, "Fra gjenstand til fortelling. Åpen arkeologi som ressurs for forskning og undervisning", in *Ringer i vann. Fleksibel læring – Kvalitetsreformen i praksis*, ed. Susanne Kjekshus Koch. (Oslo: University of Oslo, 2006), 41–54.

Ingvild S. Andreassen and Idunn Sem, "Blogg og wiki – kikkhull inn i arkeologisk feltarbeid", in *Ringer i vann. Fleksibel læring – Fem år med fleksibel læring ved UiO*, ed. Susanne Kjekshus Koch. (Oslo: University of Oslo, 2007), 75–91.

5 CONCLUDING REMARKS

Norwegian national museum databases are already on the internet, but the presentation and further development should be made carefully to avoid the pitfalls we have outlined here. The databases have become accessible for both researchers, students, cultural heritage officers and the general public. We have decided to present the original texts. These texts, written for museum purposes, will be read by a very diverse audience.

In the 1990ies, *DokPro* presented a model for the humanistic information system. Databases for a number of different disciplines were interlinked. The links within and the inroads to the events in the system were through Words, Time and Place. These are the same inroads that are chosen for the forthcoming *MUSIT* web pages at www.unimus.no. The finds are georeferenced using different accuracies. The finds are placed in time, using broad and narrow time spans. Work is carried out on ontology and thesauri. There is a need of wide classes for general and thematic searches, mid-level grouping, and as much agreement as possible between the museums on terminology for single artefacts.

One important task is to describe and sytematize the research traditions at each museum in Norway in order to facilitate searches in all collections simultaneously. Descriptions of how the information in the databases has been gathered should be readily on hand as well as digital manuals. In this way we open up our collections and can build front ends applicated to different user groups. The *MUSIT* database gives totally new venues of describing and presenting archaeological material. To explore these new possibilities and not just continue the same routines with new tools is a challenge.

In all aspects, an extensive database system is an invaluable tool as a starting point. But it should be remembered that a database is a source that should be treated as critically as any other. It is a secondary source compared with the artefacts themselves. All users need an understanding of the quality of the data. The database systems can lead the research away from the primary sources. We

think it is necessary to consider how the presentations and availability shape our understanding of archaeology and the prehistory. The presentations should be inviting, and the primary sources should be accessible for those who wish make up their own mind about the conclusions and stories presented to them.

Bibliography

Andreassen, Ingvild S. "Fra gjenstand til fortelling. Åpen arkeologi som ressurs for forskning og undervisning". In *Ringer i vann. Fleksibel læring – Kvalitetsreformen i praksis,* edited by Susanne Kjekshus Koch, 41–54. Oslo: University of Oslo, 2006.

Andreassen, Ingvild S. and Idunn Sem. "Blogg og wiki – kikkhull inn i arkeologisk feltarbeid". In *Ringer i vann. Fleksibel læring – Fem år med fleksibel læring ved UiO,* edited by Susanne Kjekshus Koch, 75–91. Oslo: University of Oslo, 2007.

Berg, Evy. " Using a GIS-based Database as a Platform for Cultural Heritage Management of Sites and Monuments in Norway" In *Digital Discovery. Exploring New Frontiers in Human Heritage. CAA2006. Computer Applications and Quantitative Methods in Archaeology*, edited by Jeffrey T. Clark and Emily M. Hagemeister, 321–327. Budapest : Archaeolingua, 2007

Hougen, Bjørn. "Tilvekstfortegnelsene. Arkeologiens Diplomatarium Norvegicum." *Universitetets Oldsaksamling. Årbok 1945–1948*:VII-XII. Oslo: Universitetets Oldsaksamling, 1950.

Holmen, Jon & Espen Uleberg. "Getting the most out of it - SGML-encoding of archaeological texts". Paper given at CAA'96 Iasi, Romania, 1996. http://www.dokpro.uio.no/engelsk/text/getting_most_out_of_it.html

Bjerck, Hein Bjartmann. "On holistic knowledge – reflections on the occasion of Anders Hagen's 80th birthday". *Primitive Tider* (2002):161–168.

Hodne, Bjarne. "Dokumentasjonsprosjektet: historikk, målsetting og utfordringer". In *Fra skuff til skjerm. Om universitetenes databaser for språk og kultur,* edited by Knut Aukrust and Bjarne Hodne, 13–18. Oslo: Universitetsforlaget, 1998.

Ore, Christian-Emil Smith. "Hvordan lage databaser for språk- og kulturfag" In *Fra skuff til skjerm. Om universitetenes databaser for språk og kultur*, edited by Knut Aukrust & Bjarne Hodne, 29–57. Oslo: Universitetsforlaget, 1998.

Solli, Brit. "From test via sign and text to things? Or: Pimp my site" *Primitive Tider* 10 (2007/08):155–164.

Østmo, Einar. "Arkeologiens arkiver på data", In *Fra skuff til skjerm. Om universitetenes databaser for språk og kultur,* edited by Knut Aukrust and Bjarne Hodne, 58–79. Oslo: Universitetsforlaget, 1998.

Østmo, Einar. Register over trykte tilvekster av norske oldsaker. *Universitetets Oldsaksamling. Årbok 1997/1998:* 177–190. Oslo: Universitetets Oldsaksamling, 1999.