

## From hand-written archive to computer-readable data

Gitte Høy

*The Central Cultural-Historical Archive, Department of Prehistory, National Museum, Copenhagen*

### 27.1 Introduction

This paper describes how the Department of Danish Prehistory at the National Museum, Copenhagen, has solved the problem of transferring a handwritten archive into computer readable form.

The archive concerned, which is at the nucleus of Danish archaeological research, is called in Danish 'sognebeskrivelsen' which directly translates to 'a parish description'. It is a description of all finds and prehistoric monuments in all Danish parishes. The description also includes several maps, on which the localities are marked, together with pictures of the prehistoric monuments. This parish description was established in the middle of the 19th century, when the National Museum sent out Land Registry maps and schedules to well-informed local people to collect information about the number of prehistoric monuments and their present condition.

During the years 1873 to 1932 the National Museum undertook a proper and systematic registration of all finds and prehistoric monuments in every single parish in Denmark. This was done thus: a Museum official and a draughtsman travelled by horse and cart around from parish to parish in a certain region, and registered all sites of relevance. A short description was made of each site, its geographical position was marked on 1:20.000 on Land Registry maps. A great many sites were also either drawn or photographed.

This handwritten parish description was continued until 1982, at which stage registration was begun directly on to computer. The reason for continuing the parish description on computer was steadily growing digging activity owing to a law passed in 1969. The law in question stated—and still states—that: 'Where during excavation work there are found tombs, burial places, villages, ruins or any other fixed ancient monuments, such work shall be suspended in so far as it affects the ancient monument.' Because of the large number of notifications of finds, a uniform method of registration and a central archive was needed. Another law was passed in 1984 which directed the country's local Museums to report all finds from their respective local areas, and made such a central registration even more necessary. The National Museum's prehistoric department therefore established an office—the Central Cultural Historical Archive—which was to maintain a central register of recently received finds and sites from prehistoric and historical times by land and by sea on EDP and which was to convert the existing part of the parish description into computer-readable form.

## 27.2 The existing 'parish description'

Before I go through the transfer to the computer, I would like to describe what the handwritten parish description looks like. The archive is, as mentioned earlier, arranged as a description of the prehistoric monuments existing in every single parish in Denmark. The parishes are administrative parts of districts, which again are administrative parts of counties. On this basis a six-figure numerical code has been created of which the first two numbers denote the county, the next two the district and the last two the parish. This number code is called the 'place number'. The individual sites within a parish have been numbered consecutively. This number is called 'the parish description number'. Each and every prehistoric monument in Denmark can thus be clearly identified by a six-figure place number and a consecutive parish description number within a parish. The parish description, in detail, consists of the following:

1. A set of maps to a scale of 1:20,000 with find place of sites marked thereon and using colour and symbol codes. For example, a grave burial mound from the Stone Age is marked with a red circle. The colour red denotes Stone Age and the circle a burial mound.
2. Along with the maps, there is a description of the individual find sites arranged in place number order. The text includes references to other archives, inventory numbers, case numbers etc.
3. Illustrations in the form of drawings, watercolour paintings and and photographs of the prehistoric monuments.

## 27.3 Transfer to computer

The reorganisation of the parish description archive to the computer started with the maps. By means of the digitizer the sites which were marked on the old maps were transferred to the computer. In future it is possible to have a map drawn of any find group to the desired scale. But the most widely used today are 1:25,000 maps with UTM grid.

When the maps were finished, work was started on the much more time-consuming part of the parish description, namely, the text. The handwritten parish description's text only gave a short description of the site, so with the transfer to the computer, it was decided to collect further information in other relevant archive groups. In this way one would achieve a parish description which contained much more information than the handwritten version. The result was that the parish description now consists of five data-groups.

1. *Identification data*: i.e. the information which is used in the precise identification of a locality.
2. *Map data*: i.e. information about the marking of the locality on the map.
3. *Site data*: i.e. information about the actual site.
4. *Case data*: i.e. information about Museum files on the locality.
5. *Free text*: i.e. a descriptive text without restrictions.

The five data groups can be sub-divided into data fields.

*Identification data* consists of data fields such as place number, parish description number, place name, association of houseowners and Land Registry number. *Map data* consists of map number, UTM co-ordinates, and sign code. *Site data* consists of site type, geographical co-ordinates, length, breadth, height, and protection number. *Case data* consists of case types, personal name, date of long-term loan, and code number of institution. *Free text* consists of text lines.

There are fixed standards for all the data fields and classifications have been introduced for some of the fields.

Over and above the place number, codes have been created for the following.

1. *Classification of the sites.*
2. *Dating of the sites.*
3. *Analysis, i.e.* whether for example radiocarbon datings have been made of a locality.
4. *Reason for registration.* It can, for example, be due to farming activity, building and such like.
5. *Case types.* Here, one can get an impression of what has happened in the 'case'. A case can be explained by a train of events. Something or other has occasioned an approach to the Museum and for this reason the Museum has done something or other. It might be that a find has been reported and the Museum therefore carries out a dig. The case types are therefore mainly of a scientific and administrative form in an institution.
6. *Institution codes.* All the country's Museums have been given a code.

Some codes have also been created in connection with the marking of the maps.

1. *Map type.* The type of map (*i.e.* which scale) the point has been transferred from.
2. *Marking code.* It is stated what method has been used in marking the site *i.e.* whether a Museums official has pointed out the spot in the field or whether the co-ordinates have been obtained by the digitizing of older maps.
3. *Sign code.* By means of the sign code, the age and type of site can be specified *i.e.* inhumation grave from the Iron Age. The sign code can be read from the map as a coloured sign.

It would be too time-consuming to go through all the codes, but I would like to mention site classifications and dating codes. The site classifications include as well archaeological as historical sites. There are in all 32 main groups which are divided again into subgroups. A site classification consists of a five-figure number code. The first two figures denote a main group and the last three a subgroup. Main groups include, for example, graves, settlements, coin finds, land transport, raw material extraction etc. From the subgroups I could mention the following examples:—cremation burial, megalithic tombs, pit, building, bridge, iron, single finds. The main groups are fixed, whereas the subgroups are floating, that is to say, the same subgroup code can appear under several main groups.

Let us take an example—a 'single find'. One can come across a single find in both a burial mound and on a settlement; therefore a single find can have both code 01145, where 01 denotes

the main group *Grave* and 145 the subgroup *Single find*—but it can also have the code 03145, where 03 denotes the main group *Settlement* and 145 once again the subgroup *Single find*.

The dating codes, which consist of three letters, cover both prehistoric as well as historic datings. The first letter specifies a principal main period; the second letter specifies a sub-period within this; and the last letter specifies an even more precise date. One has the opportunity of specifying a broad, a narrower and a very precise dating. The broad dating is used if one cannot date a site closer than say, Stone Age: AXX. 'A' represents Stone Age and the two X's 'no preciser placement within the Stone Age'. If one is able to date more precisely, *i.e.* early Stone Age the first X is replaced by an 'Æ' which represents 'early'. Thus the dating code is now called AÆX. If it is possible to give a more precise dating in the early Stone Age, *i.e.* Maglemosian culture, the second 'X' is replaced by an 'M' so that the the dating code for Maglemosian culture becomes AÆM.

Every site has to go through up to 80 datafields on each case. During the last five years we have transferred 25,000 sites from the handwritten archive to computer-readable form and we have to do another 100,000 in order to succeed. It is a very time-consuming process, but in the future we are able to search on all datafields and all words in the free text instead of having only one entry, namely the place number plus the parish description number.