



IFLA
2005
OSLO

World Library and Information Congress: 71th IFLA General Conference and Council

"Libraries - A voyage of discovery"

August 14th - 18th 2005, Oslo, Norway

Conference Programme: <http://www.ifla.org/IV/ifla71/Programme.htm>

juni 10, 2005

Code Number:

094-E

Meeting:

**107 Genealogy and Local History with Geography and Map
Libraries (Part I)**

Presenting maps and other spatial information on the Internet

Peter Korsgaard

Kort & Matrikelstyrelsen (National Survey and Cadastre)
Copenhagen, Denmark

Denmark – many decentralized digital registers

Denmark being a small country has had a long tradition for central registers or at the least registers kept by decentralised state institutions. Most of these are now in the name of e-government digitised or in the process of being made digital. One of the first to be so was the Central Person Register (CPR) combining a person's name with his date and year of birth and his home address. Now a great number have been added. Most registers are not accessible without special permission, but in my field – land ownership and land use – almost all information is accessible for all people. Partly because of the wish for transparency in government, partly because of the great administrative advantages much of this information is now available through the Internet. Regarding a piece of land you can find the name of the owner, the government estimation of the value of the land and the buildings upon it, the cadastral map, the number and use of buildings on the land, whether pollution has been registered there and in a short time also the mortgage and other liabilities. If you have special permission, there is still more to see.

These registers are not situated at one place, they are scattered among and updated by several institutions. The great task then is to see that the registers can communicate with each other in such a way that the user cannot tell that he is moving from one institution to another. The National Survey and Cadastre plays the major role in the work of making spatial data work together. In fact, this is our major reason for existing, as we do not anymore make paper maps (just a few military maps), only the data to be used for map-making.

The map archives of the National Survey and Cadastre – a mixed lot

This has not always been so. We have been making naval maps since around 1650, topographical maps since 1757 and cadastral maps since 1806. Further, in the course of making cadastral maps, we collected around 1800 a lot of maps made for land economical purposes, many of them showing the villages at the time of the enclosure. In our archives we had until 2004 – when we started sending the oldest part to the National Archives – approximately 2.000 naval maps, 70.000 cadastral maps (and the like) and approximately 30.000 topographical maps. To this can be added a collection of aerial photos taken since 1945. There is several hundred thousands of these. Most of the material concerns Denmark, but the naval maps also include the Baltic and the topographical maps include the Faeroe Islands and Greenland as well as the former Danish duchies of Slesvig-Holstein.

By far the major part – say 90% – of these are unique. They are drawings, not prints. The only place where they can be seen is at our reading room in Copenhagen. This is inconvenient for many people, and the wear and tear on this material is not to the archivist's taste. Thus for some years it has been the idea of the archives to make a digital reading room.

We are in the middle of the work – no final solutions

The following is a description of some of the thoughts and experiences made through the work of preparing this and through other projects made in the meantime. I shall heavily emphasise that we are not half ways through this, but the other projects already have given reason to re-thinking. It also takes quite some time to scan the maps (we have scanned about 10.000 maps and in the course of a rather large project plan for the sake of e-government to scan most of the cadastral maps, approximately 40.000 maps, which shall be scanned within a year from now).

A former Internet project

In a European Community sponsored project some years ago we worked together with our Swedish sister organisation (Lantmäteriet) and University of Greifswald to put quite a number of maps on the Internet (our part was 3.000 maps). The search criterion here, however, was the parish, as all the maps related to a village or a parish. Thus it could easily be kept on a alpha-numerical basis (shire-parish-village). The home page has been closed down now, but it was a great success, and the other partners have by now included their part in their own home pages. In many ways it gave much experience, not only in production flow and techniques, but in particular as the users put forward a lot of questions regarding the maps. One thing fortunately was taken care of: As many genealogists would be expected to use the Danish part, it was necessary to make search options of Denmark as it is now as well as the parish situation at the time of the maps. Many genealogists had located their ancestors through the censuses or the church records – and that reflected the old state. To include the historical parish names was then important in order to make sure that they would be able to find the right maps, even though the parish no longer existed.

Internet possibilities for the surveyors

Other kinds of experience have been obtained through the scanning, indexing and geo-coding of Surveying Documents. In Denmark, surveying is a liberal occupation, but all changes in the extent of a lot has to be approved by the National Survey and Cadastre, and all the Surveying Documents since 1906 can be found here. For a part of Denmark, all the Surveying Documents can be seen on the Internet, and for Denmark as a whole any Surveying Document approved since the turn of the year can now be seen here, too. A project of scanning all Surveying Documents at least back to 1985 (several hundred thousands), indexing and geo-coding them in order to put them on the

Internet has just been approved. I shall later return to the definition of geo-coding, but it reflects the problem that neither case no. nor lot identifier is sufficient for identification, as the lot identifier may change over time. The solution for this could also be a solution for the problem of the change over time for jurisdictional boundaries (e.g. parishes) and the like.

We are not alone

Before I turn to the technical matter, I shall emphasise that in Denmark other institutions are interested in the work of spatial references. The National Archives and the Royal Library have major collections of maps. In the case of the National Archives the need for the possibility of searching of the present and historical jurisdictions – the shires, the bishoprics, the health districts and so on shall be added. As the archive material is made and registered on the basis of the jurisdictions, they are the most important search key. Other institutions are also of importance; e.g. the Name Research Institution of the University of Copenhagen has made a great register of medieval place names and the sources and transactions (mostly deeds) related hereto.

These collections mentioned above add to the complexities, but on the other hand it is important to include these in the reflections about the solutions. The ordinary user in Denmark will have a hard time to understand why similar maps are treated one way in the National Cadastre, another in the National Archives and still a third way in the Royal Library. He would hardly understand why this should affect the availability of the map or the selection of metadata regarding the map. And the combined volume of maps and information would add to the profit for the user.

The main focus – the user

Before working out a system, it is important to turn to the future user: What does he know, what does he really want? And again there may be many answers. We shall all agree that there is a need for a search engine of the maps and other spatial information of the institution(s). Also we can agree that the user has a need for meta data of the maps and the information, though we may disagree upon the extent of the meta data. But a lot of users might be better served if we had some services, e.g. selected maps with a greater functionality. On the home page of the National Cadastre and Survey of Denmark are new and old maps (made seamless) combined with the recent Address Database of Denmark. Thus you can put in an address of today and see what the place looks like today, in the 1930s or the 1890s. Or – at another place on the home page – you again can put in the address and get a map of it with a pinpoint at the exact place (See Fig. 1). On the map there is a thread you can pull to the left or the right. To the right is a map of the 1890s, to the left one of today. Thus you can scroll back and forth between the recent part and the historical part. The last service has been one of the most popular parts of the home page.



Fig. 1. Home page of the National Survey and Cadastre (www.kms.dk). Unfortunately this part is only in Danish. The search is made either by using a place name or as is the case here by writing the street name, choosing the municipality and then choosing the address no. The pinpoint indicates the address you have chosen. To the left of the tread is a recent map, to the right a map from 1898. By pulling the thread you can scroll back and forth.

Techniques – geo-coding

Turning to the technical part, I would like to start with emphasizing the definitions I use in this article: Indexing a map is to give an alpha-numerical description of some kind in order to find it through ordinary databases, e.g. parish name, name of map. Geo-coding is to give the map a spatial reference in terms of co-ordinates. These co-ordinates relate to a system fixed in advance; of this later. Indicating a co-ordinate or an area of points indicating co-ordinates can thus be the basis for a search for relevant maps.

Geo-coding of a raster file image of a map can be obtained in different ways. One way is to point out co-ordinates to some pixels within the image. In this way the other pixels automatically also obtains co-ordinates. Usually it is done by matching points of a modern geo-coded map with points of the older map. In some cases, e.g. more recent topographical maps, you may know the exact location of the corners of the map and in this way obtain such an accuracy that you after the process can have “seamless” maps, that is that you can have viewing programs handle the separate files as if they constituted a single map.

In the process of these kinds of geo-coding you also can have a spin-off, a “bounding box”, which is a little text file indicating a polygon around the map.

```
U0410000.r01 - Notepad
File Edit Format Help
2243
04089
139941.243 078395.416
139937.602 078391.775
139952.165 078126.916
139673.653 078140.569
139679.114 078407.248
139941.243 078395.416
```

Fig. 2 . A bounding box indicating a pentangle. The first two rows concern the identifier; the following columns are co-ordinates of the points of the pentangle. There are six entries, but the first and the last are the same.

This is another way of geo-coding. There are certain pros and cons regarding the two ways.

Geo-coding through coding pixels

- takes time, especially if its going to be very good
- can be difficult with older maps as
 - they have been shrinking – not always evenly
 - it may be hard to find corresponding points – I have even seen milestones that have been moved during time
- can also as a spin off make a bounding box as a separate text file
- makes it possible to use the geo-coded raster file if you have software able to handle the geo-coding
- makes it possible that you can store the geo-coding together with meta data in the raster file, thus making it possible to copy the information within the file for use outside the institution (See Fig. 5)
- increases the file size only a little, but it still might be big as raster files tend to be
- unless you make some cuts, the geo-coding of the pixels shall concern the whole raster file, so also the white frame of the map shall be treated as part of the map. In this way you shall get more “hits” than necessary when searching

Geo-coding through bounding boxes

- is more simple and in some cases can be automated
- the file created is very small, usually < 1 KB
- if two maps cover the same area, the text file containing the bounding box made for the first map can be reused for the second map
- does not change the raster file, so it still is only a raster file without geo-coding or meta data
- bounding boxes can be created from modern vector maps. Vector maps consist not of raster data, but are based on databases and consist of points, lines and surfaces in many themes, among others often jurisdictions.

In rough terms, bounding boxes are good for searching through the Internet, while if you want to make services, the other kind of geo-coding is the right one.

In this paper there is an example of such a bounding box (See Fig. 2). It is very simple – first a few codes giving the identifier and the kind of group. Then there are some sets of co-ordinates, and you will see that the first pair and the last are identical. The co-ordinates are the corners of a polygon.

When appearing on the screen at the institution, you can choose whether these bounding boxes shall be put on top of an ordinary map or be hidden underneath it. On the Internet, though, it is the decision of the web master. Normally, the number of polygons shall be too large and just make “noise” on the screen. In some cases, though, the number of “hits” in an area is so small that it is better to see whether there are any at all. Also, as is the case of the Surveying Documents (See Fig. 3), the surveyor often is able to tell which document is the right one just by seeing the shapes. Regarding the Surveying Documents, you just click on the polygon, and the raster image of the document appears. If there were many maps covering a certain area, it would be better to have a menu of all the relevant maps described by meta data and then make your pick from the menu. That is what happens in a later mentioned case (See Fig. 4),

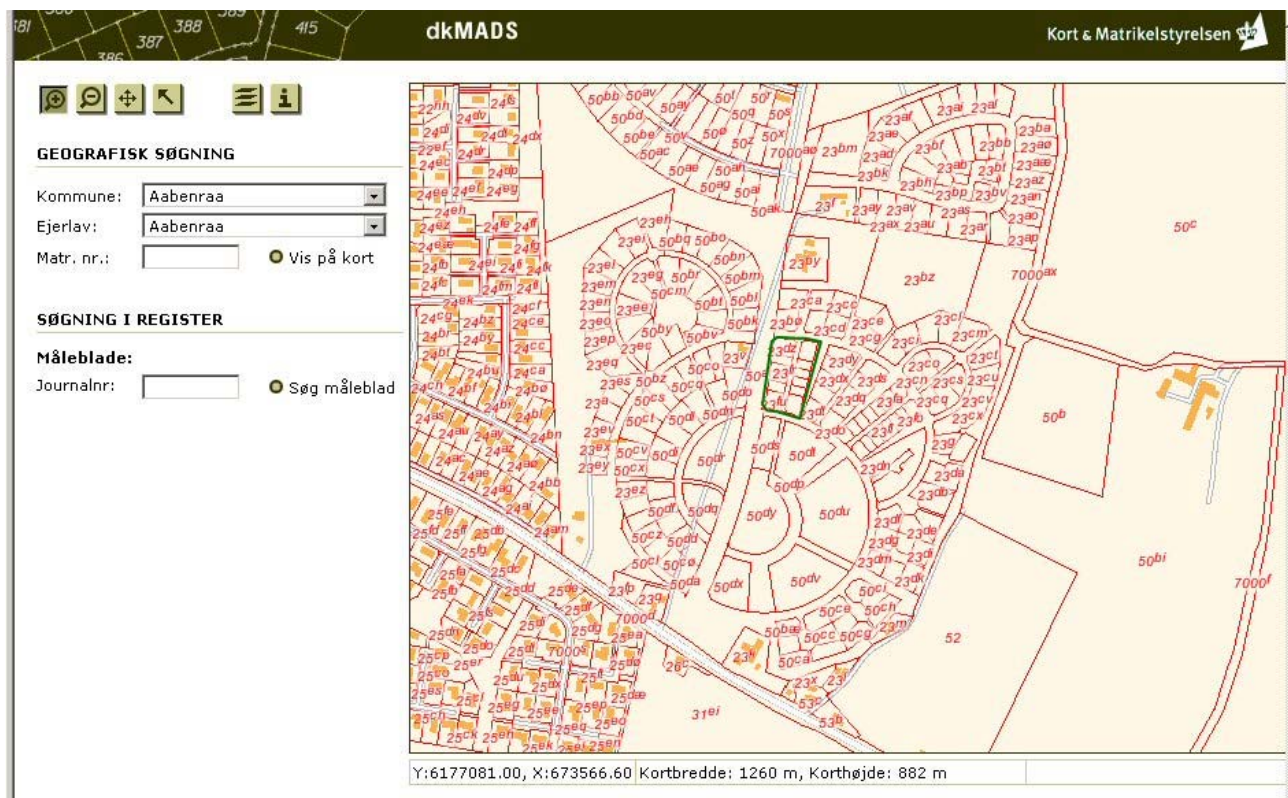


Fig. 3. Screen dump from National Survey and Cadastre – searching for recent Surveying Documents. To the left is the usual way – if you are a surveyor – by case no., parcel identifier etc. To the right is a map. There are several layers of maps, because you may need to start with the whole of Denmark and then zoom in to this layer. This is the recent cadastral map with buildings added for a better orientation. You have the borders of the lots and the cadastral nos. This map is an ordinary raster map that has been geo-coded through coding pixels. In the middle of the map is a black box, a bounding box indicating a link to the Surveying Document. By clicking on the arrow on the top left and then on the bounding box you shall get the scanned Surveying Document in a

pop up window. As still only a few Surveying Documents are in the data base, it is deemed necessary that you can instantly see whether there is a hit or not.

Re-creating old maps as vector maps could be another solution

A vector map was used by a student in order to a historical vector map of jurisdictions (villages, parishes and shires) and through this again making it possible to create bounding boxes for historical jurisdictions. Taking a modern vector map with the new boundaries and a raster map of the old boundaries and elements in 1820 did it. The last one was geo-coded and used as a wall paper underneath the new boundaries. These were then adjusted to the shape of the old boundaries and thus a vector map of the 1820 jurisdictions was created. Each little area of this map can then be turned into a bounding box and used as geo-coding element of other maps. Or, in the case of the place name index of the Institute of Names at the University of Copenhagen: the register is built around the parishes as they were about 1950 – then all the place names belonging to a certain parish could be linked to the bounding box for this particular parish at this time, giving opportunities for both seeing which place names belong to a certain parish and the other way around show the approximate location of the place name.

Searching

If you make your search through co-ordinates, you may have several options according to the software used. One common tool is a point – which is the usual for searching for addresses. In local systems this could also be extended to a circle with a user defined radius. Maybe the most common on the Internet is a rectangle, the sides going North-South and East-West. It may not always be the most efficient, but it is easy to handle for the user, especially if he himself is able to draw the rectangle with the cursor.

Another way to limit the number of hits is of course through criteria such as themes or time. The problems of this kind are so well known that I shall abstain from treating this. A source of inspiration may be <http://eu-geoportal.jrc.it/gos> . This is the home page of INSPIRE (Infrastructure for Spatial Information in Europe), an initiative launched by the European Commission among other things aiming at triggering an European spatial data structure (See Fig. 4).

The image shows the search interface of the INSPIRE European Geo-Portal. It is organized into three columns: 'Where?', 'What?', and 'When?'. Each column has a 'Help' link. The 'Where?' column features a map of Europe with country names and search options like 'Anywhere' and 'FIND A PLACE..'. The 'What?' column contains search criteria such as 'Keywords', 'Match' (Exact phrase, Any word, All words), and filters for 'Data Category', 'Provider Level', 'Scale/Coverage', and 'Type/Format'. The 'When?' column has time-based filters like 'Anytime', 'Data for time period', and 'Data updated recently'. At the bottom, there is a 'Sort results by' dropdown set to 'Content type/format' and 'SEARCH' and 'RESET' buttons.

Fig. 4. Search page of INSPIRE. As it is open for many kinds of maps, it is important to narrow down the field. To the left is the possibility of limiting the search area – you can zoom into the map.

The importance of standards

INSPIRE works for co-ordination, and this aspect shall be stressed very much. We have in the field of meta data standards such as Dublin Core and ISO 19.115: Geographic information - Metadata . In the National Survey and Cadastre we use the ISO and keep open the possibility of mapping to Dublin Core. These standards, however, do not quite cover a lot of the maps at the National Survey and Cadastre very well, as the cadastral maps have been hard to fit into these standards. In the case of the aerial photographs there is the need of more meta data, e.g. data about the camera. Besides the data of the map itself there is a need for data about the scanning and treatment of the files from downsampling to the way of geo-coding. These data are necessary if the public shall use the file. Finally there is a need for keeping up with the standards of geographic reference systems, and of course the National Survey and Cadastre is leading in the Danish field and has to keep up with the international agreements. We use the European UTM/ETRS89 (in Denmark mostly called UTM/EUREF89).

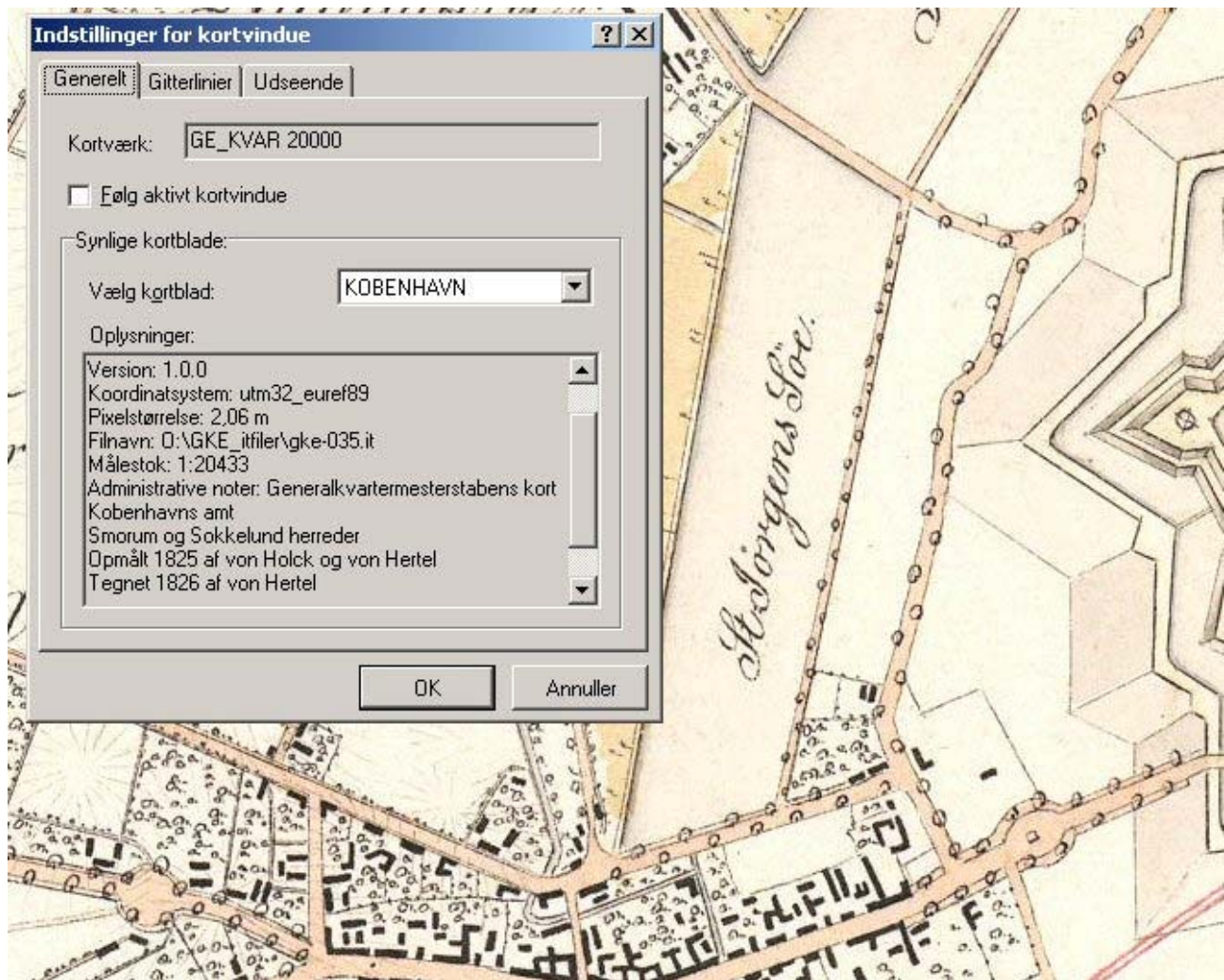


Fig. 5. An example of information stored in a geo-coded raster file. With this software, you just right-click on the map, select “properties” and then you get a pop up window telling some of the meta data of the map. However, it takes some time to write it all in the file, so you have to be a little bit choosy. But among other things it gives the name of the map series, the name of this map, the production years, and the scale. It also tells into which reference system it has been geo-coded. Otherwise it might not be very valuable to the user.

Conclusion

This has been a little of many loose ends. In a way it reflects the situation of our archive just now. It reflects of me being a historian/archivist who shall try to know some more about digital techniques. But it also reflects that there are many paths to follow, many decisions to be made, if you want go in this direction. I am quite sure this is the right direction, as we have many maps that cannot be found from the map title – and all the aerial photos on top of this. A short while ago I discussed the future way of doing this with a digital expert from my department; he just said, write down a want list and it can be done. Of course the costs of carrying it out might be to big, but it is true to that extent that it is better to have the ideas and not be limited by technical matters. I would suggest that if you think of going ahead you should at least

- describe the future users – what do they want, what are they able to do
- describe the material - what are the problems
- also think of services, not just a search machine
- do the users only need to look, or would they also like to get the files
- if you have much material, think of making it in modules to expand – but in a way that it all will fit together in the end
- think of standards of meta data and reference systems – you may later have to co-operate with others, and anyway software to transform data from one reference system to another shall be more easily at hand
- think of re-use of information, e.g. bounding boxes or data from vector maps
- find out what they do in other institutions

I am so fortunate to be in an institution that technically is quite highly advanced. Many cultural heritage institutions may not be that well endowed in this area, so co-operation among institutions may be the real key word.