ON THE IMPLEMENTATION OF ELECTRONIC CADASTRAL ARCHIVING SYSTEMS

Peter LINDBO LARSEN, Denmark

Key words: Information management; Electronic archiving; Document scanning; Best practice.

ABSTRACT

In 1998 the National Survey and Cadastre of Denmark launched a project with objective to establish a digital archive of surveying documents. Measurement sheets and sketches and other documents have been scanned and each document is described and identified by a set of metadata, which has been entered into a database. Each document is geo-coded by its bounding polygon, which facilitate the coverage of the individual documents to be visualised on a (digital) cadastral map background. A GIS-based application has been developed for distribution of surveying documents in raster formats. This application allows customers via the Internet to search, browse and to download copies of scanned measurement sheets and field plans.

Based on a case story this paper focuses on the challenges to be dealt with when launching long-term projects with substantial proportions of IT-developments. This includes the issue of operating with dynamic technology strategies that allow the project to take advantage of the latest technological achievements. Other strategic aspects as business modelling and organisation infrastructures are covered and given a practical dimension.

1. INTRODUCTION

Among the central tasks of the general cadastre is to archive the documentation of cadastral activities, e.g. registration of measurement sheets and field plans, and to make this information available for surveyors and other users. Usually, the cadastral archive is paper based, which hinder the integration of the archive into digital cadastral management systems. Furthermore, resources are allocated for time consuming archiving work by storing, looking up and making photocopies of historical surveying documents. The issue of converting the analogue archive to digital becomes more relevant as the deployment of IT supports for technical as well as administrative tasks develops. Digital archiving is an important component that brings the cadastre closer to the target of creating a complete digital cadastral data flow.

Countries or organisations who are in the phase of developing or establishing cadastral systems have a unique opportunity to by-pass analogue archives and move straight to digital storage of the cadastral documents. A digital archive is a strategic issue that ideally should be dealt with from the start of any land registration project. Because most cadastral modifications generate documents for archiving, the question of archiving technique has

severe impact on the working procedures applied for updating the regular cadastral databases. Also the frequent use of historical cadastral documents in current cadastral works means that it is worth to consider how to establish an efficient outlet of this type of documents. Digital archiving enables improvements of the cadastral administration efficiency, mainly due its contribution to a continuous and consistent data flow.

2. IMPLEMENTATION OF AN ELECTRONIC ARCHIVE OF SURVEYING DOCUMENTS (a case story)

The Danish archiving project (hereafter called MADS) covers the South Jutland County (10% of Denmark), which for historical reasons has special cadastral position. For a period of time this region was a part of Germany, and the cadastral administration was initially German in its structure. After a referendum in 1920 where people voted in favour of Danish citizenship, the region went back to Denmark, but the German structured cadastre of South Jutland remained. However the Danish cadastral system is different to the system applied in Germany. In connection with the digitisation of the Danish cadastre that started in the early 1980'es, it was decided to harmonise the South Jutland cadastral administration with the rest of the country. This digitisation/harmonisation project was completed in 1997.

Technically, the German and the Danish cadastral systems are distinct by their assignment of cadastral identifiers to land-units. Unfortunately, there does not exist one-one relationships between the two systems, which complicates the situation and a straightforward conversion is not possible.

In South Jutland the licensed surveyors request information from the archive for almost every modification of cadastral boundaries, because cadastral map updates rely on documentation on earlier surveys, which is only available from the cadastre archive. In situations where the current boundaries are documented by German structured measurements sheets, it is necessary to use the out-dated analogue German cadastral maps for searching the surveying documents. This is unsatisfactory for two reasons. Firstly, the work is time consuming and secondly, the administration has to maintain knowledge of an abandoned cadastral system.

The MADS project has provided contributions to a EU supported project under the Interreg IIC program. This project: "Uniform, Archiving and Harmonisation of Electronic Data" has 6 partners from Denmark, Holland and Germany.

3. PROJECT OBJECTIVES AND STRATEGY

The ultimate goal of the MADS project is to create a model/system for administration of surveying documents in order to by-pass the former German cadastral structure. Danish and German structured surveying documents should be supported uniformly. Users who are ignorant to the German cadastre should be able to search and to get access to German structured documents, and the archive should be maintainable by staff having no German cadastre skills.

The vision of the project is to provide users with a simple tool for searching information on earlier surveys related to particular land-units/parcels. Ideally, the information should be available to the surveyor in the field where problems occur by identifying the old boundary markers. The tool should facilitate a geographical as well as an A/N interface. The geographical interface will provide users with GIS functionality that allows them to search and select the surveying documents on a map. Quick access to the documents will be given by the A/N interface, where the documents are selected and referenced by their archive identification.

Surveying documents comprise combinations of graphical and A/N information. Evidently, the method for digitising the documents is to scan and store them as digital files in appropriate raster formats. Historical cadastral maps as raster background with current digital cadastral maps on top facilitate a solution of the problem of tracing the history of cadastral boundaries back to their German origin. This visual linkage of the two systems provides the user with information of the original cadastral identification of land-units; also the bounding polygon of each land-unit might be seen on the historical map.

Data capture (document scanning and other data conversion activities) makes out a substantial proportion of the workload presented in the MADS project. The volume of manual work involved in these activities combined with human resources available for the project made it clear that the project would last for more than 3 years. Usually it is advisable to have a well-defined target to aim at from the very start of a project. In IT dependent project like this it is not possible to predict what direction the technology will develop over a time horizon of 3 years. From the beginning the intention was to make archived surveying documents available to the public over the Internet, however, at that time the technology was not yet mature to provide solutions that would meet customers' requirements when it comes to response time. Internet tools were not powerful enough for distribution of high volume of raster data; especially, the scanned German cadastral maps with file sizes above 400 MB were a problem. Fortunately, the development of an IT-system for distribution of scanned documents was not absolutely needed until the data conversion works were completed. The strategy was to postpone the design and specification of the final distribution system to the last minute and to put the development of Internet technologies under surveillance until then.

The implementation of MADS is divided into five phases:

1. Analysis of business processes and data structure of existing documents.
2. Design of organisation and technical infrastructure and data modelling.
3. Developments of data capture applications.
4. Data capture.
5. Development of data management system for searching and distribution of surveying documents in raster formats.

4. ANALYSIS OF BUSINESS PROCESSES AND EXISTING DOCUMENT STRUCTURES

The business processes of the surveying archive were identified to be the following main processes:

1. Reception of customer's request (send by mail or fax).
2. Search of requested information in the archive.
3. Shipment of document copies to the customer.
4. Invoicing.
5. Management of archive accessions.

Without going into the detail with the processes it is obvious that more than one person is involved in processing each customer request. The time spent on searching and copying documents is not neglectable. Also the customer spends time writing the request and sketching the parcel limit of interest. Evidently, these processes could be efficiently supported by an IT-system, assuming that the documents are searchable and available on digital form. With IT support the workflow could be redesigned in a way that allows the business processes numbered 1 – 4 to merge into one process, which could be maintained by a single person. In parallel to this project the National Cadastre has launched the MIA project with objective to facilitate a digital flow of cadastral map updates from the licensed surveyor to the National Cadastre. Field plans in raster format are one of the outputs generated by the MIA system, which enables a digital transfer of surveying documents from MIA to MADS.

The analogue surveying archive is constituted of more than 10 different document types, where some types are rarely used. Of course the complexity of the archive management system increases with the number of document types to be supported. Thus it was decided to digitise only the document types that are requested most frequently. The conclusion was that three document types were selected for digitising. The volume of each document type appears in parenthesis.

1. Field plans in Danish structure (12,000).
2. Field plans in German structure (165,000).
3. Cadastral maps in German structure (5,000).

For identification and description purposes a set of metadata parameters has been defined for each of the chosen document types. These metadata will be attached as attributes to each document as the digitisation proceeds.

5. DESIGN OF ORGANISATION, TECHNICAL INFRASTRUCTURE AND DATA MODELLING

During the design phase the project components start to materialise. A successful completion of this phase is crucial, because the structures of the organisation, IT-systems and the data structures are defined/designed in this phase. Design errors might be costly to

correct in a later stage, not to mention the inconveniences of changing production procedures.

The personnel required for the project organisation should basically have skills that cover the following subjects.

1. Management.
2. Specification of IT-support systems.
3. IT-system development.
4. Data capture/conversion.

The project has been managed and staffed by internal resources. The staff has none or only little experience with development of IT-systems, but they are domain experts with knowledge of how cadastral information is structured and perhaps most important they are aware of the exceptions introduced to the general structure. The contributions provided by domain experts are very useful input for the specification and development of IT-systems. The bulk of work has been within scanning and indexing of the 180,000 field plans. The total workload was estimated by time studies.

The postponement of the design of the final data management system until clarification of the technological opportunities had severe impact on the project IT infrastructure. To leave any implementation of the management system open the strategy has been to have maximum flexibility in the architecture of the IT-support systems for data conversions. Priority has been given to open systems and proprietary data formats for data storage have been avoided. Therefore all gathered data have been stored in a central relational database management system (RDBMS), this eases the compilation of data for the final system implementation. Another pending issue is the decision on what data will be used in the final system. To minimise the risk of discharging apparently useless data that on later stage turn out to be crucial for the implementation, the strategy has been also to store intermediate data in the RDBMS. This strategy has later proven to be wise.

The analysis of the business processes resulted in a revision of the workflow. The metadata structure of the surveying documents and the workflow description are essential for the development of a business model (logical data model). Data models are useful to illustrate the complexity of the problems to be dealt with. The data model is a practical tool to describe the involved physical as well as immaterial entities and their relationships. The business model is a key design document for provision of users requirements to the IT-system. The ER-diagram presented in figure 1 illustrates the logical data model of MADS. Please observe that the model is conceptual and details are intentionally left out.

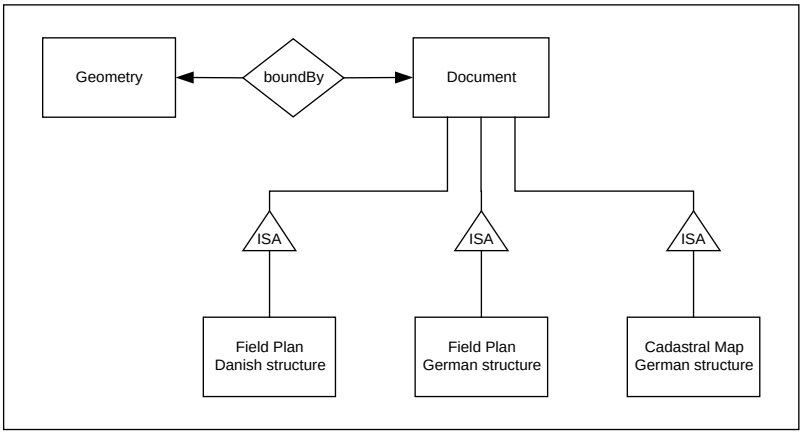

Figure 1: Conceptual data model of MADS.

The central entity of the ER-diagram is "Document" which is a super class of all documents on geographic information. The various types of surveying documents are sub-entities of "Document". The attributes attached to "Document" are those properties all documents have in common. The sub-entities are described in further detail by attributes that are characteristic to the particular document type. Each surveying document covers a geographical area of the real world. In the data model each document is allowed to have its bounding polygon presented by a list of co-ordinates in the "Geometry" entity. With this data model it is possible to identify surveying documents either by their key attributes or by their geographical footprint, which are exactly the specified selection criteria.

6. DEVELOPMENTS OF DATA CAPTURE APPLICATIONS

Several tests with scanning have been performed in order to complete the technical specification of digital surveying documents before initiating any IT-system development. The combination of age and usage of the analogue documents (the oldest are from 1870) have to some extent made the prints and drawings indistinct. For which reason the readability of scanned documents was a quality parameter that has been given high priority. The resolution of the scanning was specified to 300 dpi for all document types. The JPG data format was chosen for scanning in colour, because of its adjustable compression algorithm. Tiff was also considered but the resulting data volume would be above 1.5 TB. In 1998 disk-storage of that magnitude was almost unaffordable.

It was also considered to store the scanned data in the RDBMS as BLOBs (binary large objects). However this idea was abandoned mainly because the volume of data to store in a single database seemed overwhelming and somehow impractical to handle. The scope of the project is more in the direction of finding the best practice rather than bringing the solutions to the cutting edge of technology.

The documents to be converted to digital form during the MADS project are not only distinct by their information structure but also by other characteristics:

Characteristics Document type	Size of analogue document	Media type	BW/ colour	To scale
Field plan (German)	A5	Various	Colour	No
Field plan (Danish)	A3	Ink on plastic	BW	Yes
Cadastral map (German)	A1	Ink on paper	Colour	Yes

Five data capture tasks are identified for the data conversion:

1. Scanning of analogue documents.
2. Entry of document metadata (indexing).
3. Geo-referencing of raster data.
4. Geometric correction by re-sampling of raster data.
5. Quality checking of index data.

The data conversion work consists of 12 work packages. The table below presents the work packages assigned to each document type.

Work package Document type	Scanning	Indexing	Geo-referencing	Geometric correction	Quality checking
Field plan (German)	+	+			+
Field plan (Danish)	+	+	+		+
Cadastral map (German)	+	+	+	+	+

The volume of documents to be digitised made it worthwhile to consider software applications, which are tailored to support the specific tasks related to the conversion of each document type. The alternative of having dedicated software is to use multipurpose image processing software packages. Having tested a few types of these packages it was clear that this type of software is not very useful in a production environment, where aspects like high productivity and secure data flows are emphasised. On this background it was decided to develop dedicated IT-applications for the data conversion.

8 software applications were needed for the data conversion activities. Internal domain experts have specified the applications, and external software houses were contracted for software application developments. Although the scanning and indexing work vary with the document type, there are similarities, which to some extend allow working procedures as well as software components to be recycled. All applications are via LAN connected with the central RDBMS to which all gathered data are transferred and stored, e.g. metadata and lists of co-ordinates.

7. DATA CAPTURE

A key issue in data capture is the design and maintenance of efficient production pipes, and this is in particular due for bulk production environments. The loss of production caused by a bungling work- and/or dataflow might be considerable, because the extra time spent on producing one unit will be multiplied by total number of units. Lots of efforts have been

invested in designing and specifying software applications that support the most favourable production.

The attempts to optimise the data conversion work have provided the project with IT-tools where the operators' interferences are kept on a minimum and where data entries are guided and controlled by the system itself. Quality checks have been built into the software as well as in working procedures

Mistakes introduced by the operators cannot be eliminated totally, however, they might be minimised by building in check procedures in software and working procedures. The rate of mistakes in the data entry has been 3-5%. This makes additional quality checking necessary.

8. DEVELOPMENT OF THE FINAL DATA MANAGEMENT SYSTEM

As mentioned previously the intention of the project was to establish an Internet based service that allows customers to search, browse and to download copies of the scanned surveying documents. The overall technical hurdle for an Internet implementation was the lack of tool for extraction of information from big raster files. It was expected that the technological development would produce a solution before the project reached the phase of development of the final distribution system. Thus the introduction on the market of wavelet compressed and spatial indexed raster data formats, e.g. MrSid and ECW, were very fortunate for the project. The usual procedure to deal with raster data on the web is to transfer the entire raster file from the web server to the client, where the file is decompressed and presented in the application. This means that the response time soon become annoying even with moderate sized raster files. With the spatial indexed raster data formats it is possible to select and decompress parts of raster files, which reduces the volume of data to be processed significant. Another feature of these formats is that the raster data are divided in layers where the information of the layers is optimised for certain application scale of the data. This means that the resolution of the image matches the scale to which the image is presented. Having tested the ECW format it was clear that this technology provides a solution of distributing scanned cadastral map files. On this background it was decided to initiate the development of the final distribution system that supports German as well as Danish structured surveying documents.

The first version of the distribution system has been developed, however, the system is still in the testing phase. The client side of distribution system (see the below illustrations) is effectively a GIS-application that facilitates geographical as well as A/N interfaces for searching and looking up information in the electronic surveying archive.

Figure 2 illustrates the how geo-referenced documents (Danish structure) are visualised by their bounding polygons (appear hatched). By clicking inside one of these document footprints the user looks up the corresponding document of the polygon. The scanned documents are presented in an independent frame for further study and download.

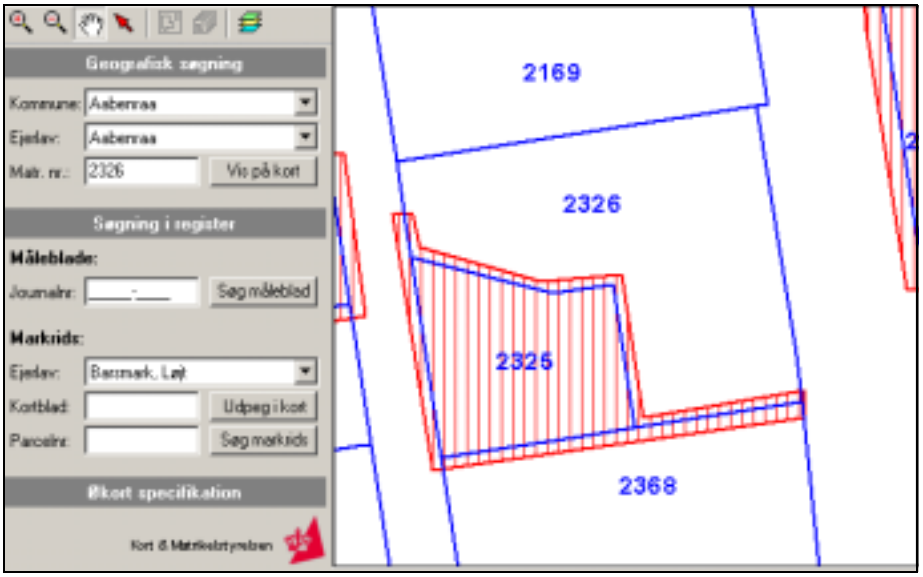

Figure 2: Bounding (hatched) polygons of Danish structured field plans.

In figure 3; the digital cadastral vector map shows the current cadastral standing, while the scanned cadastral map provides the link to the German field plans. Visually the user might identify the German cadastral number that is related to the current Danish cadastral identification. Knowing the German cadastral number the user can use the A/N interface for searching the German field plans related to the particular cadastral number.

Figure 3: Cadastral map (vector) on top of scanned cadastral map.

9. PROJECT EVALUATION

Like any other IT-system the gain of an electronic archiving system is strongly dependent on how well the system supports the workflow and how it links to the IT-systems it is

supposed to communicate with. It has been useful to analyse all business processes involving the archive, in order to identify and to describe the relevant details that should be implemented in the system. One of the lessons learned is that this type of analysis should not be contracted, because the contractor usually has limited knowledge of the practical but important details, which might lead to insufficient system specification. The workflow analysis should be based on information provided by the so-called domain experts, i.e. people who have the practical experience within their particular field of work. The benefit of involving the domain experts is dual: Firstly, the IT-system meets users requirements and secondly, the staff get accustomed to the concept of working with the electronic archiving system even before is developed and available.

Currently, the data conversion work is almost completed and the data have been compiled and stored in the final database, which constitutes the electronic surveying archive. An on-line service that makes the digital documents available to users is in the test phase. This service is planned to be in operation and accessible for the users by the end of 2001. When operational, the service will be 100% user operated, i.e. surveying documents are requested and delivered without any cadastre staff involvement.

BIOGRAPHICAL NOTES

Background

- M. Sc. in geodesy, cadastral science and planning, University of Aalborg, Denmark.
- Postgraduate studies in computer science.
- Licensed to conduct cadastral surveys in Denmark.
- Member of the Danish Association of Chartered Surveyors, Denmark

Brief career history:

1997 -	Head of Technical Department, National Survey and Cadastre, Denmark. Responsibilities: Implementation of an electronic archive of cadastral surveying documents.
1994 –1997	Seconded to MEGRIN in Paris, France, by the National Survey and Cadastre of Denmark. Responsibilities: Project manager of the Geographic Data Description Directory (GDDD), which is multinational metadata service describing available geographic information in Europe.
1993 -	Independent consultant. Responsibilities: Consultant on a World Bank financed project in Romania with objective to establish land registration systems. Giving advise on the implementation of IT-systems for cadastre and landbook.
1989 – 1993	Senior consultant, GIS-Denmark A/S, Denmark. Responsibilities: Business process analysis, data modelling and design of GIS for utility companies and local authorities.

CONTACT

Peter Lindbo Larsen
Head of Technical Department
National Survey and Cadastre, Denmark
Bjerggade 6
DK-6100 Aabenraa
DENMARK
Tel. + 45 7362 0376
Fax + 45 7462 0872
Email: pl@kms.dk
Web site: http://www.kms.dk