



Digital Historical Maps

Report from WP2

Specification of system functionality

Final version 3.6

Deliverable 9

March 22, 2001

Harry Hietanen

National Land Survey of Sweden

Table of Contents

1	Purpose of this Workpackage	1
1.1	Requirements	1
1.2	Accomplishment.....	1
1.2.1	Description of the user needs from Work package 1.....	2
2	Image production	2
2.1	Introduction and Background	2
2.2	Investigations	3
2.2.1	File Format and end products	3
2.2.2	Naming of Files	3
2.2.3	Compression.....	4
2.2.4	Geocoding.....	5
3	Storage.....	5
3.1	Database structure.....	6
3.1.1	Objectives/Aim.....	6
3.1.2	Sweden.....	6
3.1.3	Denmark.....	7
3.1.4	Germany	7
3.2	Method and Technique	7
3.3	Volumes.....	7
3.4	Browser.....	7
3.5	Security	7
3.6	Access time.....	8
4	Interface.....	8
4.1	Investigations	8
4.2	Image search system.....	8
4.2.1	Geographic search	8
4.2.2	Register search.....	9
4.2.3	Search result.....	9
4.3	Order system.....	9

Digital Historical Maps

Report	Status:	Version:	Date
WP 2	Final	3.6	2001-03-22
Systems Specification			Page ii

4.4	Delivery system	9
4.5	Billing system	9
4.6	Framework for accessing image files	10
4.7	Viewing, copying and printing image files	10
4.8	Help system	10
4.9	Information on archives	10
4.10	Information on archived material	10
4.11	Language	11
4.12	General conditions	11

Enclosures

Appendix 1	Specification of image production
Appendix 2	File naming convention
Appendix 3	Compression - Comparison between FlashPix and MrSid
Appendix 4	Compression - Comparison between DjVu and MrSid
Appendix 5	Compressing, geocoding and mosaic of historical maps
Appendix 6	Description of database – The register of historical maps and acts
Appendix 7	Database model and table structure – for Swedish data
Appendix 8	Database model and table structure – for Danish and German data
Appendix 9	Requirements and structure for Storage and Archive
Appendix 9-1	Computer System in Sweden
Appendix 9-2	Principals for Hierarchical Storage Management, HSM
Appendix 9-3	Calculated mean search time for each media type
Appendix 9-4	Draft to catalogue structure for DHM data in 3 different HSM
Appendix 10	Investigation: Search by Year

1 Purpose of this Workpackage

The purpose of this workpackage is to lay a basis for the functionality and quality to be implemented in the prototype system. The system will be based on an Internet solution comprising web pages with search paths (geographical and administrative) to the archives and to raster pictures of maps stored on a web server. The user interface includes search mechanisms, viewers, help mechanisms, information and discussion pages, ordering and billing procedures. As a proposal the raster pictures will be made available in three quality levels, with levels one and two being stored on a web server and accessible on-line.

1.1 Requirements

Requirements will be specified in the following areas for both maps and act texts:

Image production: definition of quality levels in terms of geometrical precision, colour reproduction and resolution. Definition of requirements for file format, compression method, file naming, conversion, and precision and method for geocoding.

Storage: database structure, access time, server capacity, number of simultaneous users, security procedures, linking, quality levels.

Interface: search paths and backtracking mechanisms, background maps, functions such as zoom and pan, help functions, languages, on-line service, authorisation mechanisms, ordering and billing procedures (WP5).

For the purposes of the web-based interface, existing software will be modified to the needs of the project. For image production we will use, with some modifications, the specification developed jointly by NHB and NLS. The specification will be circulated to the reference groups. Their comments will be compiled in written form and the specification modified where appropriate.

1.2 Accomplishment

Based on the report from WP1, the WP2 will develop written specifications for the end functionality of the system. The specification will be divided into three parts as indicated above (image production, storage, and interface).

The time schedule for WP2 was month 5 to 8 of the project. As the main project start was delayed about three months and WP1 report finished by the end of month 6, the work in early stage was focused on specifications for image production (WP2) and storage (WP4), to enable start of production in month 7.

The iterative work in WP2 has been integrated with developed production and storage processes. The difficulties with developing the prototype system on existing platform and demands for updated hard- and software have increased and delayed the work in the WP2 more than planned

The specifications developed in WP2 are not only valid for the prototype system, but defines also a model for managing archived data in general within the NLS. Scalability in the storage solution is

Digital Historical Maps

Report	Status:	Version:	Date
WP 2	Final	3.5	2001-03-22
Systems Specification			Page 2

therefore of great importance. Only for material from the research archive of NLS the TIFF-storage will require capacity for about 65 TB; that is 30-40 times the demand for the prototype system.

The prototype system is implemented at NLS. In the beginning an idea to have three prototypes, one in each country, was discussed. The asset of a functioning large system in the NLS and the ambition to reduce costs influenced the decision to implement the prototype only at NLS.

1.2.1 Description of the user needs from Work package 1

The user study supports functionality as follows:

Digital image files shall be of good quality; no visible information loss compare to original maps or act texts is acceptable.

Search shall be possible, both from register and a map browser.

Possibilities to see overview of the search result.

It shall be possible to look at many maps at the same time

Zoom in/out and pan functionality.

See metadata (register information) about the map.

Print out maps and act texts.

Order original files or compressed files via Internet.

Distribution of ordered files by CD, tape or downloading of compressed files.

Distributed files should be in a common usable and standard data format.

General guide for using and reading maps and acts.

Help files.

Other supported functionality:

Fast register search result.

Short delivery time on Internet.

2 Image production

2.1 Introduction and Background

These requirements and specifications aim to give an almost perfect digital reproduction from historical maps and text originals enabling further usage in presentations, research, GIS-applications, reproductions and other applications where digital information of this kind can be used. The products of maps and act texts to be delivered from the production based on these specifications are files in the commonly used and known format TIFF.

It is probably impossible to achieve a true copy of the originals; the digitising process adds faults into the copy. The ambition is to get a reproduction as close to the original as possible. Although we are

Digital Historical Maps

Report	Status:	Version:	Date
WP 2	Final	3.5	2001-03-22
Systems Specification			Page 3

trying to set up parameters and requirements that are measurable, objective and can give assumption a real high quality reproduction, where will always be demand for a visual inspection of the result in the end.

The specification for image production we will use, with some modifications, is developed by National Land Survey of Sweden (NLS). The specification is based on experiences with digital raster data since 1992.

As the goal for the prototype system is to search and view the image-files in real-time on the Internet, the files must be very hard compressed and/or presented with a technique that radically reduces the information transfer over the net.

2.2 Investigations

2.2.1 File Format and end products

Maps - File format to be used is TIFF (Tag Image File Format, ISO 12639). TIFF is the most commonly used format for raster data interchange and for lossless acquisition. Its main strengths are a highly flexible and platform-independent format that supported by numerous image-processing applications.

Text – File format to be used is TIFF Group IV. The TIFF file format with Group IV compression is used because the Act Texts are in Black & White, and other encoding formats generate significantly larger files.

See Specification for image production, Appendix 1.

2.2.2 Naming of Files

The aim of this convention is to make it easy to find out what kind of file it is. The name is therefor carrying information on the contents. There shall not be any hesitation about the files pedigree and contents.

The name is given in the production process.

The reason for a file-name convention in NLS is that, through the years, several different data capture activities have been carried through, building up different registers and storage systems. To make it possible to sort the material and build up a new, joint database, the convention for file naming is a tool.

Naming has to fulfil requirements for a logical and safe naming. Factors considered and included are i.e. map-original, text-original, geographical identity, mosaic and numbering for a single map (large map original size).

See File naming convention, Appendix 2.

See Description of database, Appendix 6

2.2.3 Compression

To be usable in a web-application, the TIFF-files must be compressed. There are different techniques and tools for compressing. The following demands have been important when choosing the tool for this project:

- The functionality in the web-browser must be fast, also on slow Internet connections (28K and 56K modem). That means that the amount of data transferred must be limited.
- It must be possible to read the detailed image information in the web-browser as well as to get an overview over the entire image.
- The image quality of the compressed file shall (visibly) correspond to the original file (TIFF) presented at scale 1:1 for an observer.

2.2.3.1 Formats

In NLS we have experiences from the following programs: LZW, FlashPix, DjVu, JPEG and MrSid. From these experiences we find MrSid to be the best tool by now to use in this type of application. The MrSid format makes it possible to fulfil the demands above at the same time as it is possible to use in standard imaging software (PhotoShop) and GIS software (ArcView, MapInfo). The software also includes functionality for viewing, printing, zooming and panning and also supports enhanced functionality in a free downloadable stand-alone viewer if you have a MrSid-file in your local environment.

As where still is an ongoing and active development in the field of image processing in common and presentation of images on the Internet in particularly the recommendation of the project group to use MrSid can be changed. The original files are to be stored in TIFF-format, and most likely usable in coming compressing tools.

It is not possible today to handle with large original TIFF files in a web browser. All original map files are too large for the application (5-250 MB each). It is necessary to compress the TIFF-files.

To fulfil the needs in the prototype several compressing programs have been tested in the project.

GIF (Graphics Interchange Formats), **DjVu** (DjVu Software), **MrSid** (Multiresolution Seamless Image Database), **FlashPix** (uses JPEG) and **JPEG** (Joint Photographers Expert Group) is some of the more used compressed formats.

GIF have a limitation in colour presentations, maximum 256.

JPEG compression is known as uneven and limited by the low resolution.

FlashPix use JPEG compression.

DjVu have high and different compression rate for each file. The software is specially designed for high quality and fast presentation of coloured and black/white documents on Internet. High compression rate decreases quality in images with many colour tones.

MrSid can compress colour images of any size, is designed for Internet and can handle co-ordinates. MrSid compression of black & white results in greyscale images.

See Compression - Comparison between FlashPix and MrSid, Appendix 3.

Digital Historical Maps

Report	Status:	Version:	Date
WP 2	Final	3.5	2001-03-22
Systems Specification			Page 5

See Compression - Comparison between DjVu and MrSid, Appendix 4.

2.2.3.2 Conclusion of compression

It's necessary to compress large original files to a more handy size and format.

Our recommendation is to use MrSID as compressing program for colour maps and act texts.

To maintain the image quality, the compression rate for the coloured maps and act texts is consequently set to near 20 times.

2.2.4 Geocoding

Almost all material from the archives concerned is, in one way or the other, geographically related, and thus possible to geocode. In a full-scale solution, digitising all documents from the archives, it would be too large an effort to bring about a complete geocoding of all documents. Therefore a compromise must be done, reducing the work, but retaining the advantages of geocoding to create fast and user-friendly functionality in search procedures.

For more detailed studies, for example when comparing historic maps with actual map data, this must be done in a GIS software.

In the project geocoding is defined in three levels, all to be carried out and tested in the project.

A-level

Overview maps with administrative division combined with a searchable register and a link to the corresponding file.

B-level

The files are combined with one pair of co-ordinates.

C-level

A complete geocoding of the entire file.

The A-level shall be implemented for all files produced. The B-level and C-level shall only be implemented as examples.

See Compression, Geocoding and Mosaic of Historical Maps images, Appendix 5.

2.2.4.1 Conclusion for geocoding

The A-level will be implemented for all files produced.

The C-level will in Sweden be implemented in a limited area due to the huge amount of maps.

3 Storage

The prototype application is proposed to be implemented as a combined geographic (map) and register search system. Depending on different types of maps there will be a separate version from each of Denmark, Germany and Sweden.

Digital Historical Maps

Report	Status:	Version:	Date
WP 2	Final	3.5	2001-03-22
Systems Specification			Page 6

The registers, one for each country, will contain all files possible to present in the web-browser. If there is a text description to the map available, this is also to be presented.

The register search is adapted to the map series involved.

The storage for more than 14000 data files accessible from a large amount of users must be based on a powerful hardware platform with good functioned software.

3.1 Database structure

To fulfil the Prototype search procedures and possibilities there should be information to search on. Each map and act have some kind of interesting information such as name, year, number, surveyor, historical events, etc.

What kind of information will we have combined to the prototype and products?

How will users search?

To manage the files the build up of a register is important. The search in register will be by geographic items. Items like County name, Municipality, Parish and Village.

See Description of Database, Appendix 6

See Database model and table structure – for Swedish data, Appendix 7

See Database model and table structure – for Danish and German data, Appendix 8

See Requirements and structure for storage, Appendix 9, 9-1 to 9-4

3.1.1 Objectives/Aim

The objectives are:

To store and keep track of the large amount of deeds/maps treasured.

To provide an easy way to search and query the database.

3.1.2 Sweden

3.1.2.1 Background

The National Land Survey of Sweden treasures a large archive of historical maps and their descriptions.

Today, the archive in Gävle contains approximately 300 000 deeds/maps.

To keep track of the large number of deeds/maps, metadata have been stored in MS Access 2.0 database. The large amount of data in the Access database has increased the response time and caused problems for multi-user use.

Within this background the database has been transferred into an Oracle database.

Digital Historical Maps

Report	Status:	Version:	Date
WP 2	Final	3.5	2001-03-22
Systems Specification			Page 7

3.1.3 Denmark

The maps (~2000) from Denmark should be registered with metadata information. There is no intention to create database only for these maps. The information is going to be implemented in the Swedish database with Danish search view.

Original files and compressed copies are storage on CD's and stored by Manage unit in KMS.

3.1.4 Germany

There is defence to build up a own database for the maps (~200) from Germany for the prototype. Metadata about the acts and maps are going to be loaded to the Swedish database.

Original files and compressed copies are storage on CD's and stored by Manage unit.

3.2 Method and Technique

The existing database in MS Access must be converted to Oracle.

The database has been created in Oracle 7.3.4.0.0 using Designer 2000. Designer 2000 were chosen as it provided the tools needed for documentation and was already existing within the National Land Survey.

3.3 Volumes

In the amount compressed files (MrSid) from Denmark, Germany and Sweden are included. The concept also shall include a solution for Swedish original files (TIFF). The original files in Denmark and Germany manage in own system.

In table below the estimated volumes for the prototype.

Type	Volume
Database	0,4 GB
Compressed files (MrSid)	50 GB
Original files (TIFF), only Swedish material	1 200 GB

3.4 Browser

The application shall be developed to work with full functionality in Internet Explorer 4.0 or higher.

3.5 Security

The prototype system shall have functionality to restrict the access to project members and members of the reference groups during the evaluation period.

It shall not in any way be possible to access original files (TIFF) from the web-application. Compressed files in the web-application only shall be readable.

Each country has restriction for own data.

Digital Historical Maps

Report	Status:	Version:	Date
WP 2	Final	3.5	2001-03-22
Systems Specification			Page 8

Mainly the security level aim to prohibit all kind of unlawful infringement in the database.
Service and lowest secure for unknown users in the prototype should be as described.

3.6 Access time

The prototype system shall meet the demands from common users, visiting the homepage

The total amount of persons who have access to the system during the evaluation period is about one hundred. A small part of them only have modem connections. The access times therefore are related to such a connection:

Search for one object in database	< 10 sec
Search for field in database table	< 20 sec
Presenting of image file on Internet (10 Mb connection)	< 4 sec
Presenting of image file on Internet (Modem connection)	< 10 sec

The prototype system shall be accessible at least 80% of the time the external web of National Land Survey is accessible and manage to meet the access times above for totally about 100 users (not simultaneous).

4 Interface

4.1 Investigations

To fulfil users needs it's necessary to start with iterative tests with Internet browsers and known technology in NLS.

Needs:

Search shall be possible, both from register and a map browser.

Possibilities to see overview of the search result.

It shall be possible to look at many maps at the same time

Zoom in/out and pan functionality.

See metadata (register information) about the map.

4.2 Image search system

4.2.1 Geographic search

This function shall make it possible to find the desired image file by zooming in an overview map. The number of levels depends on the area covered by the prototype and the actual material. The function is restricted to geographically referred material.

Digital Historical Maps

Report	Status:	Version:	Date
WP 2	Final	3.5	2001-03-22
Systems Specification			Page 9

4.2.2 Register search

The function shall make it possible to find an image file by using geographic information from the register only. In this search a successive “zooming in” shall be possible.

There are possibilities to search by Performed year but it will involve much work with corrections and developing search routines.

See Investigations: Search by Year, Appendix 10

4.2.3 Search result

As a result of Geographic or Register search a result list shall be presented. The contents in the list might differ between the organisations as well as within different map series.

4.3 Order system

The purpose of the function is to create an order of a chosen image file.

The function contain

- Choice of object
- Information on price and usufruct rights
- Choice of delivery format
- Choice of way for delivery
- Creating of customers information
- Creating of order to customers desk

In the prototype only a preliminary solution is to be implemented. This functionality has to be adapted to the environment in each organisation.

4.4 Delivery system

In the function the ordered image file is delivered to the customer

The function contents

- Creating of deliverance on CD
- Direct download of compressed file for one example from Denmark and Sweden respectively

In the prototype only a preliminary solution is to be implemented. This functionality has to be adapted to the environment in each organisation.

4.5 Billing system

The purpose of the function is to initiate invoicing. The invoice is created in regular routines of each organisation and not included in the prototype.

4.6 Framework for accessing image files

The prototype system shall be based on the conditions developed in WP5. It shall be possible to view the images in a browser free of charge. It shall also be possible for the visitor to copy and freely use the image presented on the screen.

4.7 Viewing, copying and printing image files

The aim of the function is to provide the visitor possibility of looking at, copy and print out image files. The functionality shall be good enough to judge the quality of the original or compressed file.

Functions

- Overview over the whole image
- Possibility to zoom in to a restricted part of the image
- Possibility to pan the image
- Possibility to copy selected part of the image
- Possibility to print out selected part of the image

In the final design of the functionality the opinions of the reference groups in the evaluation period shall be considered.

4.8 Help system

The function aims to help the user to manoeuvre the prototype system

Functions

- Help for navigating the geographic search
- Help for navigating the register search
- Help for navigating the Mr Sid Viewer
- Help for order image files
- Help for downloading image files

4.9 Information on archives

The function aims to give the visitor a short introduction to the archives concerned in the three organisations.

4.10 Information on archived material

The function aims to give the visitor an introduction in the material available.

Functions

- General information about material in the archives
- Detailed information about the map series in the prototype

Digital Historical Maps

Report	Status:	Version:	Date
WP 2	Final	3.5	2001-03-22
Systems Specification			Page 11

- Information about the contents and how it can be interpreted

4.11 Language

The prototype system shall be multilingual.

Functions

- General information shall be available in four languages, Danish, English, German and Swedish (for example the Help system).
- Detailed information about the material shall be available in the language of the country from which it originates and English.

As for the rest of the homepage, it will only be available in English.

4.12 General conditions

The final layout is chosen during the evaluation period, regarding the demands from the reference groups.