

The Challenge of Preserving Digital Maps

by JANI STENVALL

This paper addresses some general matters on the preservation of different types of digital materials. The well-known technical problems are explained and also some of the usual strategies to resolve these problems. Digital maps and their preservation are considered in the light of the future legal deposit act in Finland.

*“The great creator is the great eraser. .”*¹

INTRODUCTION

Before discussing the preservation of digital maps I would like to explain some general points about preserving any digital material. The wide range of different digital products and their special characteristics impose a great challenge to libraries. The file formats, software needed to render the publication, and hardware requirements are developing rapidly and this means obsolete environments for older digital materials. The fast moving computer technology is providing new updated software and “improvements” all the time. This development is leaving a trace of obsolete systems, formats and above all the documents that we used in these obsolete environments. The computer industry that is creating today’s digital world is also erasing yesterday’s digital world.

Preserving the bits, the bit stream of any digital document, is quite easy because we are able to make perfect copies of the original product (if they are not copy-protected). However bits are the same for every file format, we cannot tell whether a certain bit stream is for text, images, a few musical notes, or geo-data without interpreting the bits with the right software and environment. We cannot say that we have preserved something for the future, unless we can render and use the preserved object in the future. These aspects suggest that documentation and metadata will be a key issue when handling and administering large amounts of digital data.

In addition to metadata we will need some guidelines how to keep these publications “alive” during time. The current methods or strategies for preservation are migration and emulation. Under the concept “migration” we can put copying and conversion of the data. We might need to refresh our documents and copy them to a new medium, from CD-R to DVD-R, or we might find a certain file format unsupported and convert the document to a current file format. We migrate the data to keep it alive for future scholars. We can rely on standards even though it is known that one can find many standards for one single problem and that new standards will replace current ones.

It is to be expected that in some cases we might find that migrating the publication will lose some of the characteristics of that particular document. For example, a CD-ROM with interaction and multimedia is difficult to migrate to another environment without losing the original feel and serious effort to rewrite the source code. Emulating the older environment in the current computer platform will probably give us the same look and feel of the original CD-ROM. Emulation is the computer science approach to preservation and it might prove itself as a useful strategy. Both these major strategies have pros and cons. They are both likely to be used in a “memory organisation” which is trying to preserve digital documents. The metadata with technical, provenance, context and processing information will help us in the preservation and rendering of these documents or at least provide some backup information for the digital archaeologists of the future.

Preserving digital documents is also an organisational problem. The long-term planning is left to the libraries, archives and museums. Publishers are not so much interested in preserving their publications for the future, even though for some materials there might be a market in the future. The amount of digital data is growing so fast that it is hard to keep up with preservation. Who or what organisation should do this work, how will they have proper funding, expertise and continuity? There is always the question of what to preserve. What selection guidelines we should follow? These questions are central and one way of looking at these problems is to turn to libraries and to legal deposit.

LEGAL DEPOSIT OF DIGITAL PUBLICATIONS IN FINLAND

The revision work is still going on for the new legal deposit act in Finland. In several countries this has already been done. Hopefully we will have the new law operational by 2004. The main new concern for the Finnish legal deposit libraries is the inclusion of digital publications. The digital publications are considered to be important materials for future researchers. Off-line products as well as the Finnish on-line (Internet) publications are to be collected, catalogued, preserved and put to scholarly use inside the premises of our six legal deposit libraries. This applies of course to the different types of digital maps. The draft for the new national copyright act has some statutes that enable this depositing.²

The preparation of Finnish digital deposit legislation has been underway for some time now. The National Library (the Helsinki University Library) has some experience of web archiving, CD-ROMs, copy protections, e-books, cataloguing digital documents etc. In web archiving we are now testing special software, which is designed to download documents from the Finnish domain (.fi) sites as well as from other domains known to be Finnish. This harvesting is gathering static web pages. The first harvesting round resulted in 11,7 million separate files (ca. 700 Gb), but still less than a terabyte of data. We have

no means of automatically gathering dynamic pages or sites with restrictions (like registrations or passwords) or commercial material that you need to buy in order to use the digital document. The deep or invisible web is not harvested. The publishers are required to deposit these “hidden” publications (with accompanying metadata) to the deposit libraries. Depositing is also used for all types of off-line products. This legal deposit will result in different digital data archives, which requires careful planning. We are also now acquiring a Digital Object Management System to govern possibly all these different types of digital publications.

VISIONS OF DIGITAL MAP PRESERVATION AND USAGE IN THE HELSINKI UNIVERSITY LIBRARY

Digital maps are one of the publication groups to be regarded also as legal deposit materials. At the moment the solutions to archive and preserve digital maps are being processed. The following is still based on my personal speculation on how digital map preservation and services in the Helsinki University Library might develop and come into being.

We can put digital maps into few categories, such as:

- maps published on CD-ROM or on other off-line media
- simple on-line maps (as image files, digitised map collections)
- special map files (commercial products sold as image files)
- on-line map services (interactive, database driven, GIS products)

These categories come from my own, non-expert view, on digital mapping. I understand that there are many other mapping systems and spatial data available, but I limit my paper to “published” digital map products.

Preserving digital maps

Essentially large amounts of digital maps can be defined as legal deposit material. The basic methods for creating a national collection of digital maps are automated harvesting and depositing by the publisher (or manufacturer) as mentioned earlier. For preservation purposes the library needs the digital maps with metadata and without any copy protection or encryption. The mass of digital maps is great and, considering the practical work, it seems that we need to select and prioritise the maps we are going to preserve, and to put for scholarly use. Maps that have an important national or scholarly function should be the ones that we are providing in the first place. Full production and complete digital map collections will follow.

The Challenge of Preserving Digital Maps

We still have the basic map products available in print form. So we are not necessarily in danger of losing all cartographic materials of our time, as paper is still much better than a computer file when considering longevity. However the draft of our new law is saying that parallel publications (e.g. manifestation in both print and file) should both be deposited to the library.

Many digital products, not just maps, are compiled or made on-demand. The customer orders the tailored digital document from the content provider and this ordered product might be the only one ever made. Clearly, depositing cannot be done for all the possible variations of the product made available by the publisher. It seems that “normal” publications that are predefined and on the shelf should be regarded as legal deposit materials. But in the future we may well face the situation where everything is produced on-demand. If we want to preserve the historical information through maps we should have deposited copies of some selected on-demand products. The on-demand model is related to large map databases. Map or GIS databases as a whole are hard to be fully deposited in a library and maintained. We should make a clear distinction about databases and their contents. According to the UNESCO Guidelines for Legal Deposit Legislation³ the databases that have raw data (that is unorganised data that could be selected and gathered on order by an individual to create an object) should be regarded out of scope for the legal deposit. This would mean that organisations acting as legal deposit organisations should collect only the database materials that are independent, separate and complete units (such as text or image). The nature of the data in dynamic or database driven services defines whether it needs to be deposited or not. For example, the Topographic Database (Maastotietokanta) produced by the National Land Survey of Finland could be deposited to our system as separate files. One file would be one image derived from the database. If we choose the scale to 1:10,000 for the whole of Finland this would mean about 14,000 files. This collection of images could be updated once a year; each year only the map areas that have been updated could be deposited. But if we follow the UNESCO guidelines this database would be considered out of the scope of legal deposit. Even though the end-product is a basic raster image, the data in the database are unorganised, and the contents of one map image is selected on demand by the customer (in this case the library). So far we do not have a clear policy regarding map databases and the on-demand output from these databases. These are essential questions that need to be addressed in the legal deposit arguments before the law will come into force.⁴

CD-ROMs or similar off-line map products include usually their own software providing the user the environment for map viewing. These are off-line products and emulation should be regarded as the main strategy for these publications. The program that is included in the interactive CD-ROMs is hard to migrate and we probably cannot preserve the “look and feel” without emulating the old environment on the current platform. This means copying the data out from the disc and storing it intact somewhere. An emulator

package designed for the current platform together with the stored data will give us the opportunity to use the products.

Map producers also sell digital maps as separate products (raster or vector images). Without considering again the on-demand dilemma, these products could be deposited by the producer/publisher to the library. Migration seems an appropriate strategy for these types of maps. For digital map files we could try to find some migration guidelines, to convert them to current file formats every once in a while. The different map file formats need to be investigated and possible standards should be followed.

Simple on-line maps (which we will get through harvesting), contain all the image files used to depict maps (GIF, TIFF, JPRG, PNG, etc.). The files are archived and that is it, but we could also decide a specific moment in time when all, let us say GIF-files should migrated to a new file format. We will not produce manual metadata for these Internet image files because of the large quantity of data: we are gathering the whole Finnish Internet without any special selection guidelines. Interactive web maps that require some input from the user cannot be harvested as such. Dynamic pages are out of reach from the current harvesting software. In this case we are interested in deposits of “snapshots” of digital map images from the service, not the dynamic map service as a whole. For some dynamic map services the viewing of map images might be so tightly bundled with the service environment that it may be hard and expensive to extract the map publications from the map service. This needs more studying.

Through this legal deposit act we are now building a large collection of different types of digital maps: on-line images, CD-ROM products and digital map objects in different file formats. For preservation we need to have detailed metadata of these objects and some general guidelines on what type of procedures we are doing to keep these digital maps alive. We do not want to have a graveyard of digital maps in obsolete environments. Metadata is one of the key issues. The metadata scheme or set needs to be selected and tested. Hopefully the metadata standards used to describe digital maps can be used at least partially for preservation and in information retrieval purposes as well.

The access and use of digital maps

Once the digital maps are archived and catalogued (or indexed somehow), the researchers and other users are able to use them through dedicated workstations located in the premises on Finnish deposit libraries. The Copyright Act will restrict the usage of these archived materials and careful security measures will follow. The workstations need to have proper software installed to enable map viewing. Map images harvested from the Internet are likely to be common image file formats, but special map file formats (e.g. MapInfo) need specific viewing software.

The Challenge of Preserving Digital Maps

Cataloguing can be done in the Digital Object Management System (DOMS) using the metadata available as basis. The metadata in the DOMS can serve different functions: bibliographical descriptions, searching and retrieving, use conditions, preservation etc. The DOMS should have connections to the national bibliography (the Fennica database) and allow records to be copied and converted to MARC-format. The Fennica database will hold information on selected digital maps. We are also considering collection level descriptions for special map image collections. These bibliographic databases are searchable (with possibly one integrated interface) for the researcher to find relevant maps.

The possibility of using deposited digital maps is of course an important service for the researchers. This affects the library customer services, which needs well-trained map curators and other personnel to know all these materials and to assist the researchers.

THE ROAD AHEAD

This paper gives some general ideas on what is planned and what still needs to be discussed regarding Finnish legal deposit and digital map services in the Helsinki University Library, but we also know that the actual practical work is demanding. We need co-operation with the publishers to learn about different digital products. We need to prepare all the legal deposit libraries and ourselves for these new services. In different technical matters and in preservation work the devil is in the details.

We are just now realising the complex tasks involved in preserving digital materials, including digital maps. As you all know maps are very much historical documents, visualisations of our surroundings and an important source materials to many types of future researchers. The future map scholars might need the help of map curators and digital archaeologist to be able to display and use some digital maps as they were used in the early 21st century. This is truly a challenge – to avoid the loss of digital publications and keep our cultural digital heritage alive.⁵

NOTES

1. Brand, S. *The Clock of the Long Now - time and responsibility*. London : Phoenix, 1999.
2. Hakkarainen, J. "The Reform of the Legal Deposit Act in Finland". *Helsinki University Library Bulletin*, 2002.

3. Larivière, J. *Guidelines for Legal Deposit Legislation*. A revised, enlarged and updated edition of the 1981 publication by Dr. Jean Lunn, IFLA Committee on Cataloguing. Paris : United Nations Educational, Scientific and Cultural Organization, 2000. <http://www.ifla.org/VII/s1/gnl/legaldepl.htm>
4. Mattila, R. *The Legal Deposit System in Finland: From Printed Matter Towards Electronic Publications*. http://www.nlib.ee/inglise/textid/conf_info/2sund.html
5. Rothenberg, J. "Ensuring the Longevity of Digital Documents". *Scientific American* 272 (1995) 1, 24-29

WEB SITES REFERRED TO IN THE TEXT

FENNICA - the Finnish National Bibliography:
<http://www.lib.helsinki.fi/english/libraries/fennica.htm>

National Land Survey of Finland. <http://www.nls.fi>

PADI – Preserving Access to Digital Information. <http://www-prod.nla.gov.au/padi/>