CO-265

PRESERVATION IN DIGITAL CARTOGRAPHY

JOBST M.
Vienna University of Technology, VIENNA, AUSTRIA

ABSTRACT

The preservation of digital cartography may lead to the digital cartographic heritage in future. One main requirement is the understanding of structural, organizational, legal and technological aspects.

1. INTRODUCTION

The working area of cartographic heritage more and more uses digital technologies in order to sustainably keep digital cartographic content. A high dependency of technologies to time is only one single aspect of cartographic heritage. The ongoing usage of geospatial services, productive recording methods and their incorporation in digital cartography that tries to integrate users´ recordings in combination with remote sensing and digital cartography show up futher aspects of a complex preservation process.

2. DIGITAL CARTOGRAPHIC HERITAGE'S COMPLEXITY

Modern cartography is heavily influenced by digital approaches. Reproduction processes as well as dissemination procedures make use of digital mechanisms that mostly enhance traditional processes. This technical changes lead to new and extended applications. In terms of cartographic heritage these technological developments of neogeography result in new challenges for enabling sustainable cartographic heritage for the future. Their conceptual structure, intermedial dependencies and technical requirements lead to a more complicated framework.

The characteristics of modern cartography and its paradigms lead to a strong interlacement of all used components that have to be born in mind if sustainable archiving should be done. The following depiction of a conceptual cartographic heritage architecture shows main dependencies of considered core components. On one hand the triangle in the graphic additionally shows the grade of digitalization, on the other hand cartographic heritage depth can be defined. The grade of digitalization starts with content and its storage media. In principle this categorization begins with analogue media, like a paper map on its storage media paper and a printed content. As soon as a digital content has to be processed, the device and format of the data become important. Consequently the processing application and its dependencies have to be considered. Cartographic heritage depth covers content-based- and artistic-based parts of historic values. Thus a very first description of cartographic heritage starts with storage media, its material, fabrication and condition. In a next step the content with its syntax, pragmatics and semantics adds to storage media. Storage media and content form an artifact, which allows to suggest the resulting application and usability framework. Finally full cartographic heritage depth for an digital cartographic tool additionally covers device, format, application and most of all the interface of the map.

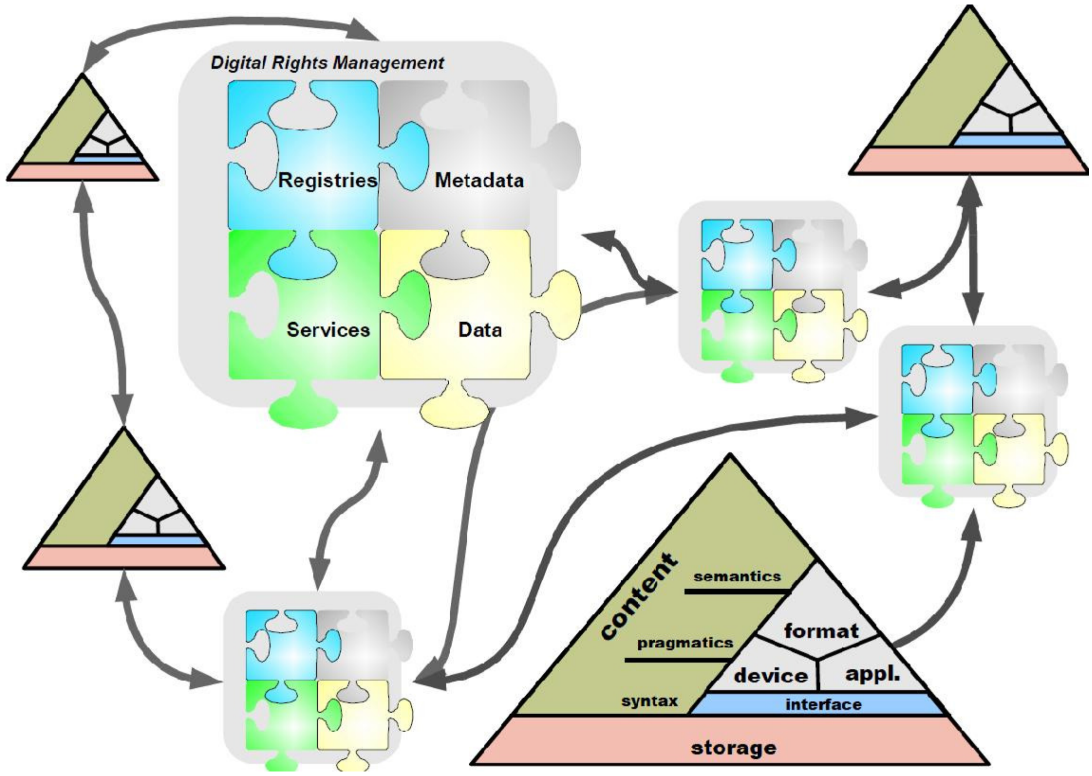

Fig.1: An overall structure of digital cartographic heritage including the technological base of Service-Oriented Architectures.

2.1. Principle components of digital cartography

The main components in digital cartography from an archiving viewpoint form a fundament for cartographic heritage. If one of the components cannot be successfully archived, the access to the whole cartographic application is in danger. Thus these components are closely related to each other. The components can be classified to content, format, application, device and storage, which all together imply specific strategies for sustainable archiving.

2.1.1. Content

The content of a cartographic application is the information media. This means that the selection of information according to a context or map use and the graphical coding of this selection builds up the content. For this reason semiotic rules have to be considered, which helps to make content understandable and usable [Chandler 2002]. Semiotic rules cover semantics, syntax and pragmatics. In terms of cartographic heritage it becomes important to clearly associate semantics, syntax and pragmatics of a content to its operational area. Then its statement and aimed application becomes clear and can be used for alternate use or proceeding interpretation. In general a picky selection of metadata facilitate powerful search algorithms for the content or provide additional background knowledge of the content, like metadata of information acquisition and precision will inform about data quality.

2.1.2. Format

The format describes the structure of the digital document, which should be read/used by an application. For cartographic heritage the documentation of this structure is inevitable important, because this will help to build an application in case of inaccessibility. In addition to the format documentation, the kind of format is of importance. Two various kinds of format can be identified: binary and ascii. While a binary format is a direct machine code, which cannot be humanly read, the ascii format can directly read with any text editor. Thus the ascii format should be favored in terms of cartographic heritage. Format standards like Extensible Markup Languages (XML) are also of the form ascii. Therefore the format GML as XML specialty gains also importance for preservation issues.

2.1.3. Application

The application forms the processing part of a multimedia map, which is some kind of software. Generally an application calls for specific hardware requirements, like the playing of sound requires a soundcard and speakers or virtual 3D environments need graphic cards, appropriate drivers and specific displays. Therefore the application is closely related to the device.

Similar to the format, an application can be encapsulated in a compiled proprietary format or readable as open source application. Regularly the software has to be compiled for reason of operating system and processing unit usage. A proprietary application, which is not well documented and lacks of an API (Application Programmable Interface) or similar programmable extensions, can hardly be adapted to new operating system- and hardware environments. If the source code of an application is accessible, this piece of software can also be adapted to any new hardware environment. A lot of examples of various Open Source initiatives prove that an open source code enables individual extensions and adaptations to operating systems and hardware environments, like the Open GIS software GRASS [www.grass.itc.it], which runs on any operating system, even on PDA's, in the meantime.

2.1.4. Device

The device is the interface between computer and human being. This interface plays a central role for an expressive and effective transmission of information. For cartographic applications the interface plays the role of information carrier. The interface makes geospatial information accessible for the human sensual system. Depending on the graphical resolution of the interface, geospatial information depth of the map has to be adapted/prepared in order to keep the content perceptible. If a specific interface does not exist anymore, the original information content cannot be transmitted again. For example stereoscopic information for lenticular displays on a standard display will show up illegible fractals due to the splitting of information according to the lenticular lenses. If this specific information should be sustainably archived, also the interface is needed for a correct presentation.

2.1.5. Storage

Storage is the most important aspect for cartographic heritage, as it is for any heritage topic. As soon as an element can be accessed in future, it can be called heritage. For digital cartographic heritage with an increasing amount of data, information, applications and so forth, new storage procedures have to be developed. The old paradigm of keeping and saving is only partly true for digital media, which also request constant temperature and atmosphere. But most of digital media's lifetime does not exceed 20- 30 years [Borghoff 2006]. Therefore these kind of media call for a constant copying to newer storage media until a next migration period has been reached. Within these processes of copying mistakes in writing the content and the table of contents have to be avoided. As soon as the table of content does not fit the data any more, the archive is destroyed.

The archiving components of digital cartography can be considered in an isolated way, each component for itself. In fact this isolated approach would be not sufficient, because each component shows up relations with the others. Furthermore the use of the Internet and distributed data and services impose pressure upon archiving strategies. These technologies can be subsumed in Service-Oriented Architectures (SOA) for further descriptions.

2.2. SOA-based structures

The archiving of a SOA structures does not only concern technological means. In point of fact the responsibility and process methodology for the protection of geodata and maps has to be discussed in order to structure a technological framework. Assume a working archiving framework, like specific Services in the Internet, and no one who is responsible for parts of the cartographic heritage structure (content, meaning, format standards, ...), then the digital content and entire sequences within the SOA network will definitely be lost. This loss can be exemplarily observed within the picture database flickr (www.flickr.com), when images are removed and links become outdated. In order to overcome this barrier of cartographic heritage, main important aspects that exist today and are needed for a prospective archiving of SOA, have to be identified.

SOA characteristics. The concept of Service-Oriented Architecture (SOA) requires loose coupling of services with operating systems and other application-based technologies. Loose coupling means that classes of programming have almost no knowledge of each other. Beside a clear definition of interfaces, operations and attributes of classes are exchanged and adapted for the coupling in real-time. Services maintain a relationship that minimizes dependencies and only requires that they maintain an awareness of each other. The coupling, or interfacing/binding, of SOA services generally makes use of XML, though this is not required [Bell 2010].

Following the aim that services should be bound for an application, those services have to be found in the network. Therefore established services have to be published, which means that their existence becomes maintained in online registries and their capabilities and descriptions stored in accessible meta-databases. This basic principle of SOA leads to its main components: services, metadata and registries. Furthermore services may access data. So these resources (data) have to be online and accessible as well.

From this point of view a lot of questions arise when thinking on service-oriented architectures: How can we archive service-oriented applications, that depend on the Internet, communication protocols and ad-hoc connections? The conceptual architecture may help to keep the main dependencies within the range of vision before going into more detail.

3. ACTUAL PRESERVATION STRATEGIES

Preservation strategies vary for applications, formats, content semantics or devices [Borghoff et all 2003]. Instead of one single technique to comprehensively save cartographic heritage, different methods are needed for the different parts. Generally three main strategies for digital long-term preservation can be distinguished: migration, emulation and the computer museum.

3.1. Migration

According to the final report of the task force on archiving of digital information [TFADI 1996], migration is the periodic transfer of digital materials from one hardware/software configuration to another or from one generation of computer technology to a subsequent generation. The main purpose of the migration is to preserve digital objects´ integrity in a way that clients retrieve, display or use these objects with changing technologies. The main four aspects of migration are replication, refreshment, repackaging and transformation [Lavoie 2004].

One main organizational measure that has to be considered by archives is the restriction of format diversity. Migration steps increase with diversity of formats. Additionally last migration steps should be kept for a valid fall-back strategy. This aspect at minimum doubles the size of the archive [Borghoff 2006].

3.2. Emulation

A complete different approach to preserve digital objects is emulation, which conserves the native environment of a comprehensive system. Emulation can generally be adapted to the levels of application, operating system or hardware [Granger 2000]. Additionally the concept of virtual computers is a derivation from the emulation concepts. A virtual computer delivers access interfaces to standard devices, which are implemented at a software level that runs on a hosting operating system (which may change during time) [Foster and Kesselman 2003].

The main advantage of emulation is that all objects within the emulation, thus within archives, do not need modification. The only modification has to be adapted to the emulator, which should work on actual operating systems.

3.3. Computermuseum

The approach of the computermuseum tries to provide original hardware and software systems for at least a medium range of time. As long as original parts exist, hardware defects can be repaired cost-effective. From such time as no original parts exist, alternative parts have to be used or original parts have to be reproduced, which increases the costs for keeping a system online dramatically [Krebs 2011].

For a long-term archiving, computermuseums are important components, because it is the only one that can provide a reliable authenticity of media in- and output [Borghoff 2006]. The authenticity of original output is the most important for cartographic heritage, because cartographic information is always adapted to the output media/information carrier.

3.4. SOA preservation approaches

There are a lot of initiatives for archiving the WWW. These web-archiving initiatives collect portions of the WWW and store these parts in archives. Web-crawlers are used for an automated collection because of the massive size [Brown 2006, Brügger 2005]. Although all links are stored in the archive, these copies miss the dynamic characteristic of the Internet and can only document one single state of time.

Other approaches make use of SOA for digital preservation. These approaches build up single modules that are needed within digital preservation, like format conversion- or migration services. At least the preservation architecture can be accessed by individuals as well as archives at the application layer [Ferreira 2007] in order to fulfill preservation tasks.

Digital cartography that relies on SOA heavily uses a corporate network or the Internet. This framework does not use SOA modules for preservation processes, but calls for preserving the SOA structure, interfaces and contents. This means that a mapping of the network structure or at least its functionality, protocols, actuality and integrity needs to be solved.

4. ONE EXAMPLE OF FUTURE-ORIENTED PRESERVATION INITIATIVE - EXPANDING THE BASICS

The complexity of the digital domain leads to the change of paradigm “store and save” to “keep-it-online”, which initially helps to expand the basic understanding for digital preservation. This expansion is a technological and procedural one. It focuses on the SOA principle with its distributed data that are connected via services across standardized interfaces.

Disregarding cartographic applications for the first step, the preservation of digital geospatial data encompasses a range of content types that reach from raster data to vector, databases or other types. Each of these types poses numerous preservation challenges. Geospatial data does not consist of one specific classification, but often consists of combinations of types within specific software environments [McGarva 2009].

One main aspect within the preservation of digital geospatial data is the establishment of international standards. It comes along with standards for geospatial data exchange and usability, which are crucial for geo-enabling the Internet or wireless and location-based services. A main organisation beside industrial standards, that are build by ISO, is the Open Geospatial Consortium (OGC). The OGC is an international industry consortium of over 400 companies, government agencies and universities that work together on publicly available geospatial standards. Examples of this work are Web Mapping Service, Web Feature Service, Geography Markup Language or the OGC KML [OGC 2010a].

Most of OGC activities focus on service-oriented scenarios. Therefore inserting a temporal component into these services will be important to enable consistency and interoperability across time. There is the vision that the development of geospatial archives can be supported by web services in a more efficient and automated way [Morris 2011].

The data preservation working group of OGC was initially formed to address technical and institutional challenges posed by data preservation. One main goal of the working group has been the creation and dialog with the geospatial community and archival community, which have to stake in data preservation. Further goals of this working group are the identification of intersections between data preservation issues and OGC standard efforts, as well as the introduction of temporal data management use cases into OGC discussions [OGC 2010 b].

One possible way to preserve complex end-user representations as documents in terms of data preservation is via GeoPDF that captures and embeds interactions with services as well as outputs from GIS desktop software environments.

Although the standardization initiatives expand the basics of geospatial preservation in terms of interfaces and data formats, these considerations do not cover cartographic multimedia applications or interactive Internet maps, which consist of much more “components” than the data component. The complex framework of preserving digital cartography leads to actual main foci that represent pragmatic steps in solving long-term availability of digital cartographic heritage.

5. MAIN FOCI OF PRESERVATION IN DIGITAL CARTOGRAPHY

Nowadays the field of archiving digital cartography is not in the first focus within the geospatial domain. Instead new applications that support geospatial access, -exchange and enable user-participation grow up. A lot of applications make use of historical geoinformation, which has to be available in appropriate quality. For some reasons “historic“ digital maps, like CD-Rom atlases or similar, may be accessed to extract their interpretations of geoinformation. In many cases a loss of this knowledge will be observed due to missing components (cartographic heritage structure) of CD-Rom applications.

The preservation of digital cartography may also show up additional values of the digital geoinformation and its semantic interpretation within cartographic applications. These values are hard to define beforehand. They rather occur incidentally. Actually the borders for digital cartographic heritage seem not to be on a technological basis, but on organizational and legal structures. Actual working topics cover access to cartographic heritage, quality aspects in reproduction, interface and format standardization, intellectual property rights as well as organizational and legal aspects. On one hand it is important to provide easy access to cartographic heritage, which includes analogue and digital access, in appropriate quality, on the other hand the development of standards, organizational and legal issues open the doors for sustainable access/usage of cartographic heritage.

5.1. User-friendly access and dissemination

Dissemination is of central importance for digital cartographic heritage. The reason is that dissemination builds up the knowledge on existence of an increasing amount of digital cartographic artifacts. An efficient extraction comes along with specific search algorithms and selection methods. In the end the user interface delivers user-friendly access to the search and access of geospatial artifacts.

5.2. Issues of digitalization quality

In conjunction with dissemination and cartographic heritage the digital reproduction of historic maps is an central activity. Often the digitalization is referred to preservation of the analogue content without reflecting a change of archiving paradigm. Doubtless the main advantage of digitalization is an easy access to artifacts. But a possible replacement of the originals depends on the reproduction quality, which definitely has to be 1:1. This means that the digital artifact has a resolution as high as the printout would need to gain the same quality as the original. Information of data carrier characteristics (paper specifics), that may be lost in digitalization, can generally not be stored in the digital copy. These issues have to consider file sizes, archival file formats and quality- and process management methods.

5.3. Development of Standards

At the moment the development of standards focuses on the data and service level. An archival standardization at the application level cannot be considered at the moment due to technical and organizational restrictions.

Starting from the experience of PDF/A, the archival profile of PDF, a discussion has started to create an archival profile for GML. GML is the main important standard for geospatial data exchange. Although GML appears to be a promising solution for data preservation, there are some complicating factors. One important is that GML as specialty of the XML is not a single format, but covers several community specific implementations (add-ons) that are embedded as profiles within GML versions. In order to lower the barrier of implementation the OGC released the Simple Features Profile in 2006, which is a constrained set of GML [Vretanos 2005] and a starting point for an archiving profile in GML.

5.4. Organizational and legal aspects

Open questions regarding organizational and legal topics arise when a possible way to embed digital cartographic heritage in archives or libraries as well as the open access to cartographic heritage in map libraries shall be processed.

For the first issue, findings of case studies within International Research on Permanent Authentic Records in Electronic Systems (InterPARES 2) brought up the actual scientific discussion on organizational aspects in cybercartographic preservation [Lauriault 2010]. The case studies based on the “Cybercartographic Atlas of Antarctica” and “Preservation Practices of Scientific Data Portals”. It could be shown that the greatest challenges limiting the long-term preservation of an Cybercartographic Atlas are neither technological nor procedural. Instead it is the fact that archives and libraries are not in the position to ingest cybercartographic products, because they have no resource capacities to do so. Resource capacities could be adapted if the legal framework and key procedures would be established.

For the second issue, the open access to cartographic heritage in digital map libraries, it becomes obvious that an approach to the Intellectual Property landscape is needed. Therefore shared notions, conventions and practices have to be defined. As soon as digital cartographic heritage is available via Spatial Data Infrastructures (SDI), its distribution/usage is connected to digital rights management (DRM). The DRM bases upon Intellectual Property Rights (IPR) of a provider (library) and concerns the temporality of the information as well as the temporality of the IPR based on national laws [Fernández-Wyttenbach 2011].

This section showed impressively that even if we can solve technological and procedural topics of digital cartographic heritage´s preservation today, we still have to make some efforts in organizational and legal issues. From a temporal perspective these issues need much more time and consensus of involved parties than the technical ones.

6. CONCLUSION

This contribution illuminated the complexity of digital cartographic heritage, showed up actual preservation strategies and discussed main foci of preservation in digital cartography. It could be shown that preservation of digital cartography goes beyond digitalization and the change of paradigm from “store and save” to “keep-it-online”. Especially Service Oriented Architectures call for organizational, legal and technological frameworks that will help to sustainably keep digital maps accessible. A first step could be done by standardization of exchange formats and thus creating an archiving GML profile. Other perspectives may lead to the preservation of complex end-user representations as data-documents via GeoPDF. Both approaches will only help for the archiving of data components. Therefore the work on preservation of digital cartography needs a wider approach that reaches from an “organizational and legal consensus” to the “development of preservation standards for all cartographic heritage´s components”, “minimum requirements for digitalization/reproduction quality” and “user-friendly search and access”.

7. ACKNOWLEDGMENTS

The author would like to acknowledge the work of all participants in the book project “Preservation in Digital Cartography: Archiving Aspects”, which was published in 2010 at Springer, New York. All participants brought the discussion in this important cartographic working field forward. The book is an

unique reference for the actual state of preservation in digital cartography. A lot of components are still not considered and call for further investigation.

8. REFERENCES

Bell, M. (2010) SOA Modeling Patterns for Service-Oriented Discovery and Analysis. Wiley & Sons. pp. 390. ISBN 978-0470481974.

Borghoff U.M., Rödig P., Scheffcyzk J., Schmitz L. (2006) Langzeitarchivierung – Methoden zur Erhaltung digitaler Dokumente; dpunkt Verlag; Heidelberg.

Brown, A. (2006) Archiving Websites: a practical guide for information management professionals. London: Facet Publishing. ISBN 1-85604-553-6.

Brügger, N. (2005) Archiving Websites. General Considerations and Strategies. Aarhus: The Centre for Internet Research. ISBN 87-990507-0-6. http://www.cfi.au.dk/en/publications/cfi.

Chandler Daniel (2002) Semiotics – the basics; Routledge Taylor&Francis Group: New York. ISBN 0-415-35111-1.

Fernández-Wyttenbach A. (2011) An Intellectual Property Rights Approach, in Jobst M. (ed.) Preservation in Digital Cartography, Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg 2011.

Ferreira D. R. (2007) Business Networking with Web Services: supporting the full life cycle of business collaborations, in Pallab Saha (ed.), Handbook of Enterprise Systems Architecture in Practice, pp. 419-433, Information Science Reference, IGI Global.

Foster I., Kesselman C. (2003) The Grid. Blueprint for a New Computing Infrastructure, Morgan Kaufmann.

Granger S. (2000) Emulation as a Digital Preservation Strategy, D-Lib Magazine. Vol.6, No.10.

Krebs N., Borghoff U.M. (2011) State-of-the-Art Survey of Long Term Archiving – Strategies in the Context of Geo-Data / Cartographic Heritage, in Jobst M. (ed.) Preservation in Digital Cartography, Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg 2011.

Lauriault T. P. (2011) The Preservation and Archiving of Geospatial Digital Data, in Jobst M. (ed.) Preservation in Digital Cartography, Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg 2011.

Lavoie B. F. (2004) The Open Archival Information System Reference Model: Introductory Guide. DPC Technology Watch Report 04-01.

McGarva G., Morris S., Janée G. (2009) Technology Watch Report: Preserving Geospatial Data. Available at: http://www.ngda.org/docs/Pub_McGarva_DPC_09.pdf (accessed June, 25th 2010)

Morris S. (2011) Preservation of Geospatial Data, in Jobst M. (ed.) Preservation in Digital Cartography, Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg 2011.

OGC (2010a) Open Geospatial Consortium, 2010; OGC Abstract Specifications, available at: http://www.opengeospatial.org/standards/as (accessed June, 25th 2010).

OGC (2010b) Open Geospatial Consortium, 2010; OGC Data Preservation Working Group, available at: http://www.opengeospatial.org/projects/groups/geormwg (accessed June, 25th 2010).

TFADI (1996) Preserving Digital Information, Report of the Task Force on Archiving of Digital Information, The Commission on Preservation an Access, Washington DC.

Vretanos P. (2005) OpenGIS Web Feature Service (WFS) Implementation Specification. Open Geospatial Consortium, available at: http://portal.opengeospatial.org/files/?artifact_id=8339 (accessed June, 25th 2010).