

Archiving in Digital Cartography and Geoinformation

Berlin, December 4/5, 2008

- Report on the Workshop -

Markus **Jobst**, Potsdam/Vienna⁽¹⁾; Horst **Kremers**, Berlin⁽²⁾;
Eunice **Gill**, Richmond⁽³⁾; Jozef **Hernik**, Kraków⁽⁴⁾; Zdenek **Stachon**, Brno⁽⁵⁾

(1) Hasso Plattner Institute at University of Potsdam markus@jobstmedia.at (2) CODATA-Germany, Berlin office@horst-kremers.de

(3) National Archives, Richmond, U.K. Eunice.Gill@nationalarchives.gsi.gov.uk

(4) University of Agriculture in Kraków, Poland, Faculty of Environmental Engineering and Land Surveying rmhernik@cyf-kr.edu.pl

(5) Masaryk University Brno, Czech Rep., Laboratory on Geoinformatics and Cartography zstachon@geogr.muni.cz

The workshop „Archiving in Digital Cartography and Geoinformation“, which was held from December 4th to 5th 2008 at the State Archive Berlin, was visited by an International and interdisciplinary group of experts. A total of 66 participants were registered. 19 contributions were presented over two days. To some extent the presentations could be supported by actual applications and professional solutions within the exhibition area.

The workshop was made possible by the patronage of the German Cartographic Society of Berlin-Brandenburg (DGfK) in cooperation with CODATA-Germany (German National Committee for CODATA, Committee on Data for Science and Technology of the International Council of Science), Hasso-Plattner Institute of University of Potsdam (HPI), the Commission on Digital Technologies in Cartographic Heritage of the International Cartographic Association (ICA) and the Berlin State Archive.

International and national experts from the areas of cartography, geoinformation, geosciences and technology were invited for participation. The aim was to discuss long-term storage, archiving methods and best archiving practices in response to the growing demand, as experienced by many participants in their daily working life. This demand on archiving concerned disciplines of cartography, GIScience, archiving and technologies in order to work on actual archiving methods, database archives, technological-, organizational- and legal requirements for a sustainable future for cartographic heritage.

Workshop Program and Structure

In the course of the workshop the interests of international and national participants were considered. Thus the conference language of the first day was English and of the second day German.

Day 1

After a short address of welcome by the organizer, chair of the German Cartographic Society Berlin-Brandenburg Branch and chair of CODATA-Germany, Mr Horst Kremers, the director of the Berlin State Archive (Landesarchiv), Prof.Dr. Uwe Schaper, and co-chair and representative of the

HPI and ICA, Dr. Markus Jobst, the program was explained and the organisers informed participants to hold active discussion at the end of each of the conference days.

The first contribution with the title „OAIS for cartographers“ presented the standard OAIS (Open Archival Information System) as fundamental concept for archiving geoinformation. Central problems within cartographic contents, which occur with carrier media change, especially with vector-based data, were exemplarily shown: damage of metainformation and reference frames, change of colour codings or displacement of names.

The second presentation with the title „Dependencies in Neo-Cartographic Heritage“ focused on modern technological developments in cartography and their proposed archiving in order to enable accessibility to cartographic heritage in the future. Complex archiving structures result from digitalization, multimedia content and Service-Oriented, whose contents cannot be sustainably stored by current procedures and technologies.

The third contribution with the title „Map Archive of Institute of Geography“ elaborated on map archives of the Institute of Geography, Masaryk University in Brno. This contribution described planned digitalization process of the map archive as well as design and standards of metadata, file formats and georeferencing.

The fourth presentation given by The National Archives of the UK with the title „Heritage Inspires“ focused on existing spatial data infrastructures in Europe. Examples showing the temporary nature of information on the Internet depicted analogies and references for service oriented architectures of geodata. Especially change of borders, time scales and prospective historical atlases of Europe were a main focus.

The fifth contribution with the title „Archiving, Management and Dissemination of Geospatial Data and Maps“ exemplified requirements for an open, scaleable and components based system architecture in backup and tape systems with a hint on „geospatial intelligence“ (GEOINT).

The sixth contribution with the title “Archiving the Complex Information Systems of Cultural Landscape for Interdisciplinary Permanent Access – Development of Concepts” reported on the complexity of information systems for archiving cultural landscapes in Europe, which was a main result in an EU Interreg III project called “Cultural Landscapes”.

The seventh contribution with the title “Pitfalls in Preserving Geoinformation – Lessons from the Swiss National Park” demonstrated difficulties in long-term data series with meta-meta-database systems, changing data models and storage of massive data layers by means of best practices of the Swiss National Park.

The eighth contribution with the title “Preservation Partnerships for State and Local Government Digital Geospatial Data in North Carolina” reported on the data archiving project NCGDAP from

North Carolina, which focused especially on time-based data management, long-term storage of geospatial data and the role of spatial data infrastructures for archiving.

The ninth contribution with the title “Archiving and deriving georeferenced spatial information from historical mine maps” referred to an economic use of historical maps and mine drawings and described the preparation of historical mine information.

Day 2

The tenth presentation with the title “Großformatdigitalisierung – Erfahrungen und Schlussfolgerungen aus Projekten” depicted the organization of scanning procedures, digital processing and embedding of meta-information.

The eleventh contribution with the title “PDF/A for scanned and digital maps” referred to the development of the format PDF as archiving standard for static documents. The planning for prospective PDF developments was presented as well as improved compression algorithms and applied methods.

The twelfth contribution with the title “SIARD – eine neue Lösung für die Archivierung von relationalen Datenbanken” introduced an initiative of the Swiss National State Archive, which expands and changes the archive to digital business administration data. Therefore archiving methods for relational databases with the help of the new introduced standard SIARD were presented. Similar approaches for geospatial content are subject to pilot studies at the moment.

The thirteenth presentation with the title “Bereitstellung und Nutzung historischer Luftbildbestände im Landesarchiv NRW” described how the approach of Hansa Luftbild was created and the assembling of the aerial image collection of the Main Archive Düsseldorf (Hauptarchiv).

The fourteenth presentation with the title “Digitalisierung und Verwaltung von 1,5 Mio. Katasterdokumenten” reported on digitalization, data management and the software solutions that were used.

The fifteenth presentation with the title “Positionspapier der DGfK Kommission Angewandte Kartographie” focused on a composition of guidelines for long-term storage of digital data, especially for the collection and archiving of geospatial- and graphical data, which were developed by the project group “digitalization of ancient maps”.

The sixteenth contribution with the title “Rissarchivlösungen für die Katasterverwaltung” reported on the successful conversion to CAS technologies of the historical cadaster archive and its accessibility with high performance.

The seventeenth contribution with the title “Geodatenbezug für Laien und Experten” described results gained in student projects on the pragmatic access to geospatial data and aerial images with server-based methods.

The eighteenth contribution with the title “Archivierung digitaler Kataster- und Topographieinformationen” mentioned a number of various archiving strategies and legal appointments for archives to preserve maps (Hessisches Archivgesetz).

Both days ended in extensive discussions and round tables of all participants, which clearly stated the need of an interdisciplinary dialog between experts. This dialog is fundamental to any coordinated interdisciplinary developments in the complex domains of archiving in digital cartography and geoinformation. Contributions from various disciplines play an essential role for the future of cartographic heritage. Therefore, topics and motivations for further action were collected:

Topics and Motivations for further Action

The contributions of the archiving workshop reached from practical approaches to theoretical concepts, which were mixed with detailed and active discussions. This mixture led to an active incorporation of all participants, whose occupational background finally demonstrated needs and prospective developments.

Protection of Data

The discussion on protection of geospatial data was based on the conclusion that open formats (including appropriate norms and standards) have to be used. Nevertheless it is likely that a massive loss of data will occur due to human factors (lack of knowledge), missing descriptions, software-based and hardware-based damages.

Human factors

Human factors, especially lack of knowledge, predominantly concerns the relocation of interest. For instance a dataset becomes declared as unimportant because these data are not used for actual tasks anymore or the data are beyond the field of activities. Then the elimination of this dataset is a next step that is mostly justified by a saving on resources. Prospective uses of these datasets and their future importance (in order to answer prospective questions) is generally not considered.

A possible solution for human factors: the precise regulation of roles/responsibilities.

Missing descriptions terms and definitions (Metainformation)

Missing descriptions in terms of geospatial data lead to the main loss of temporal and task-oriented reference as well as loss of geospatial references (geodetic date, reference systems, specific offsets, ...). In general geospatial data becomes unusable and is lost to further applications.

A possible solution for missing descriptions: the establishment of standards for metadata and meta-metadata, which allow for embedding and direct linking inside the original data formats. Important is the specification of extension of some mandatory core with elements demanded by map archiving community which consists of cartographer, informatics, archiving specialist etc.

In addition appropriate search algorithms are needed.

It is also important to unify terminology for archiving process (e.g. OAIS). This will be helpful for minimizing of interdisciplinary human factor induced errors.

Software-based damages

Software-based damages (anomalies) lead to “sneaking” deformations of original data. For example the copying of data or transformation to another data format may lead to a different data structure and thus also to a change of coding or even changing description of geodetic reference (in the internal mathematical description of data). It is obvious that this imperceptive procedure is not reversible. It mainly causes a pragmatismal loss of the dataset.

A possible solution to avoid software-based damages: usage and establishment of possibly open standardized data formats in order to avoid transformation to new formats.

Hardware-based damages

Hardware-based damages/anomalies/deterioration often lead to unreadable data carriers (storage media) and thus a loss of all contained data/information.

A possible solution: Constant copying to new data carriers considering technical requirements in terms of compatibility and their expected lifetime.

Difference of „analogue-born“ and „digital-born“

The determination whether geospatial data existed analogically and were digitized afterwards (analogue-born) or were digitally created and do not have any tangible original (digital born) has to be made. Whereas analog-born data can follow the predominant paradigm of archiving (store and

save), digital-born data call for new methods in archiving. The reason is that analogue-born data can be created by the original anytime, often with much higher quality due to improvements in digitizing technologies. Digital-born data do not have an original master, which can be used for “digitalization”. Instead concepts for long-term preservation are needed in order to access these digital originals.

One possible attitude for quality preserving digital data, can be represented by storing original data together with their visualization rules.

Digitalization of analogue data enables an easy dissemination, access and professional analysis. Just the step of digitalization cannot be seen as “archiving” of the original material! In fact the original is needed for a renewal of the digital representation. In contrary a reconstruction of an original from a digital representation calls for appropriate qualities: in case of maps, which mainly consist of graphics and line art, reproduction specifications will need at least 1200 dpi and lossless data compression.

Geospatial data and -applications, that only exist as digital-born artwork, lead to new archiving strategies, like for instance migration or emulation. These methods try to keep digital born content read- and accessible.

Selection of Geoinformation and Maps according to „Long-Term Value“

The growing amount of geospatial data and applications leads to the question of “archiving appropriateness”: What kind of content should be saved? What criteria are used for content selection? Who is responsible for selection (archivist, librarian, information scientist, cartographer,...)?

A marked discrepancy can be observed between professional aspects for archiving depending on archivist, librarian and geodata producer. The understanding on the “long-term value” varies with the professional background. Whereas a selection of geospatial basic data is difficult to define because these data form the basis for answers to various (not yet defined) questions, a selection of geospatial applications and products seem to be useful. In this context it becomes obvious that the definition of “long-term value” has to be resolved. Geospatial products provide a definite reference number (e.g. ISBN) according to their date- and place of publication and publisher. For this reason these products can easily be catalogued and put into a library. Of course the technical effort required for archiving digital products needs to be evaluated (with interdisciplinary assistance). The role of Digital Object Identifiers (DOI) has to be evaluated in all phases of archiving in digital cartography and geoinformation. An exclusion from the catalogue may be associated with the use of open standards, open source and appropriate data carrier for the specific geospatial product. In addition an inclusion of digital maps into the library’s catalogue calls for the technical equipment to store the digital content in an accessible state. Equally legal issues

for geospatial content restrict their incorporation in libraries: the free and public access to geospatial data depends on copyright laws and access may be possible after 30 to 70 years (depending on the national legal framework).

Complexity of Information Portfolios

Several contributions showed future requirements for the analysis of cultural landscapes and long-term availability of information inventories quite plainly: the complexity of geospatial datasets is growing exponentially. Comparisons with homogenous datasets, like those resulting from scanning, clearly show additional requirements as not only is the file-size of datasets growing, but also interrelations of features and attributes concerning e.g. cultural landscapes have become important. Additionally the access methodology to these complex structures need further investigation. The UNESCO convention for intangible cultural heritage (<http://www.unesco.org/culture/ich/>) explains the importance of a sustainable handling of nature, universe and cultural developed techniques. Herein the indemnification of interoperability, consistent archives and ad-hoc analysis of complex applications are main foci for further methodical developments.

Summary of Deficits, Recommendations and Next Actions

The main result of the workshop and discussions with all present professions was that there is an urgent to make long-term availability and archiving of geoinformation and cartography a main subject of interdisciplinary and public discussion. Main demands for clarification are a clear legal situation, assigned competencies and manuals for best practices (as wiki website or in the form of books). At the moment the minimum for preserving cartographic heritage can be depicted by the motto “keep it simple and online”, where geodata should be kept as simple as possible (standardized and additionally as image information) and “online” describes the actual state of knowledge about geodata, their metadata and meta-metadata. This actual state has to be kept alive with changing technical requirements in order to keep the access to these data. The big aim is that archiving in scientific, legal, cultural and economic areas should be self-evident within the frameworks of processing infrastructure.

In order to go on with the fruitful discussion and define next branding milestones, the following steps and next meetings are planned:

- To enable and keep up an intense and interdisciplinary communication a discussion forum (eMail) was established. Further information and registration can be found at <http://www.kbx7.de/?lid=18531&c=list>.
- All authors of archiving workshop's contributions and members of the commission on digital technologies for cartographic heritage of the ICA were invited to participate a book project on the prospective cartographic heritage, which will be printed by Elsevier or Springer. Further information and deadlines can be found at: <http://www.cartography.at/heritage/>.
- The results of virtual discussions and latest developments should be posted in the Internet. Therefore a "blog" website is in preparation, which will be found at „<http://blog.cartography.at>“ in near future.
- We invite interested groups and experts to report on her/his own field of activity and provide these reports for the community. A pertinent collection of these activity reports will then be published as "best practices" , e.g. best practice for scanning ancient maps, data management or similar, as reviewed e-journal with the title "Archiving Perspectives in Geoinformation and Cartography".
- A workshop of the ICA Commission on Digital Technologies for Cartographic Heritage will be held on April 6th and 7th 2009 in Venice
<http://web.auth.gr/xeee/ICA-Heritage/Commission/workshops.htm>
- A workshop on Archiving is in preparation for Berlin: http://www.codata-germany.org/Archiving_2010/
- Archived accessible materials can create useful input help for establishing of process of ubiquitous mapping in the sense of described by ICA commission on Ubiquitous mapping.
<http://ubimap.net>
- Promotion of the importance of digital cartographic heritage to the governmental institutions and also to the wide public.