CO-267

MAP ARCHIVE ACCESSIBILITY USING THESAURI

STACHON Z., KOZEL J.
Masaryk University, BRNO, CZECH REPUBLIC

1. INTRODUCTION

Research related to spatiotemporal changes is always based on geographic spatiotemporal data. Today, spatiotemporal data is usually stored in computer databases or in another suitable file formats. However, for searching and finding appropriate dataset among many others, the high quality metadata and catalogue service are necessary. In the field of geography, there exist well-known standardised solutions based on OGC Catalogue Service [13], ISO 19115 Metadata [6], and ISO 19139 Metadata XML Schema Implementation [7]. Generally, geographic metadata record consists of author, space, time, and attribute information, so metadata querying supports searching by author, space, time, or attribute filters, or its combination.

On the other hand, in pre computer era, spatiotemporal geographic data was stored in paper maps or in paper books and tables. For research related to spatiotemporal changes of geographic phenomena, old maps are often crucial and sometimes even the only resource of geographic phenomena historical state. Compiling spatial data from more old maps regarding its time horizon, spatiotemporal changes of monitored geographic phenomenon can be observed and described. Furthermore, content of old maps can be important for actual planning, documentation and cartographic applications (Jobst and Gartner [8]). However, a usage of such information is limited by possibility of finding and accessibility.

For searching old maps, usually bibliographic retrieval service is used. Each map is represented by one bibliographic record, which should be encoded according to one of existing bibliographic standard such as MARC 21 [11] or DublinCore [3]. Each bibliographic record, similar to geographic metadata record, consists of author, space, time, and attribute information. Attribute information in case of old maps usually describes main thematic focus of the map and it is represented by keywords e.g. “geology, iron ores deposits”. However, keywords do not usually describe all types of objects displayed in maps (i.e. map content). Furthermore, meaning of keywords is quite vague unless they refer to existing thesaurus (or code list). To solve these two issues, we present an idea of Old Map Object Thesaurus and associated Old Map Indices Thesaurus. Research is made on old maps from map archive of Institute of Geography, Masaryk University.

2. MAP ARCHIVE OF INSTITUTE OF GEOGRAPHY MASARYK UNIVERSITY

Archiving of maps, atlases and other cartographic items has last on the Institute of Geography of Masaryk University almost century. Nowadays, there are about 18 000 inventory items. Except of separate map sheets there are the map sets, atlases, globes etc. Maximum amount map archive achieved at the end of the Nineteen Seventies. Later few thousands of map sheets had to be passed on military topographic service due to security reasons. Fortunately mostly only newest maps sheets had to be passed on. At present days is number of items in the map archive increased only slowly by casual purchases.

Oldest prints in the map archive come from sixteenth century (e.g. first atlases of Abraham Ortelius etc.). The map archive is specialized on maps of Moravia, which collection is unique. Also maps of Bohemia and other countries can be found in the map Archive. Maps come from most of famous historic cartographic workshops in Nederland’s, Germany, France, England, etc. There is traditional production of facsimile of chosen maps for wide public and experts.

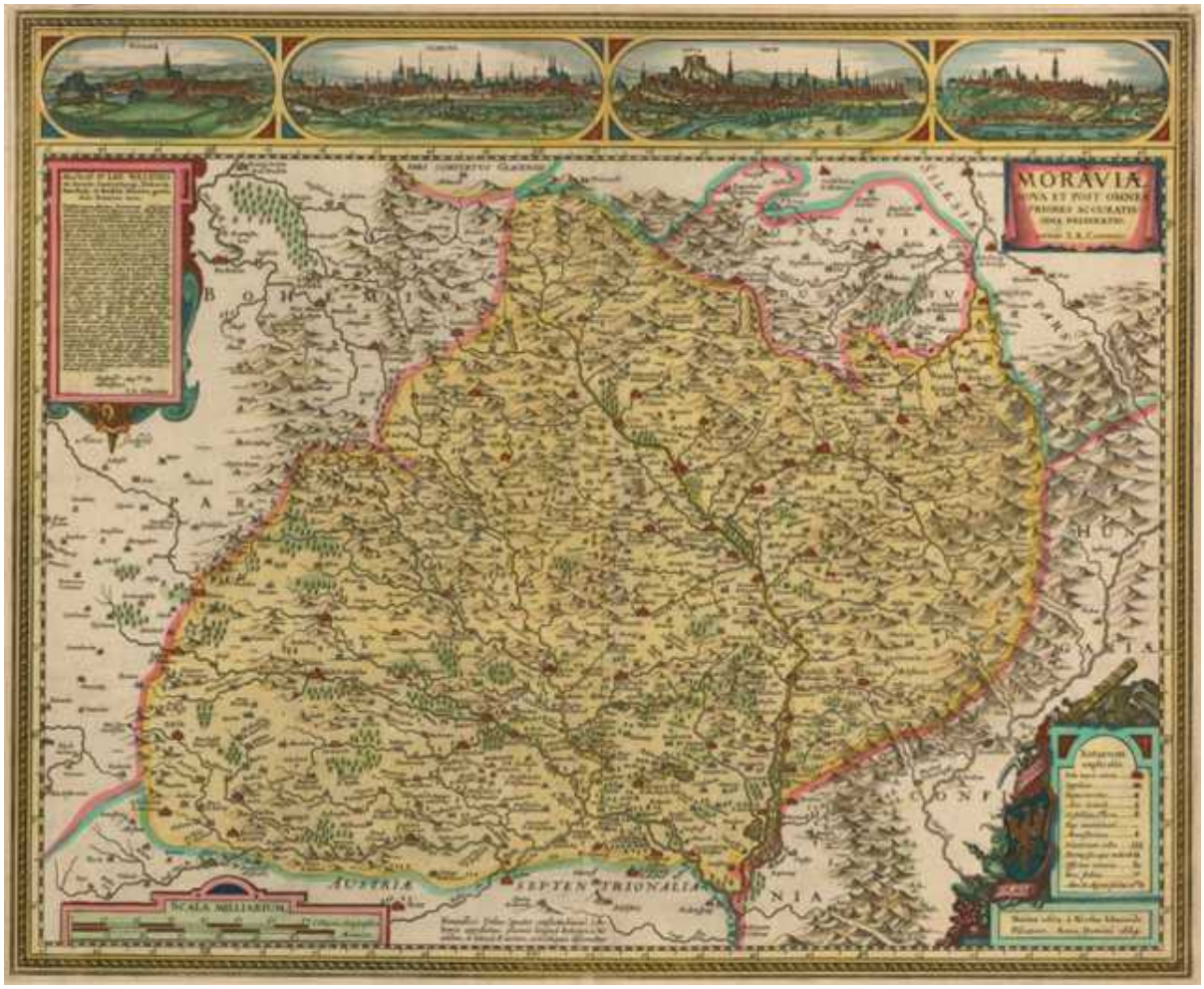

Figure.1. Comenius map of Moravia from map archive (KMMA, 1627)

The end of the Twentieth and the beginning of Twenty-first century brought new technologies and also change of concept of old maps storing, accessibility and also usage. It can be said, that technological development enabled extension of aesthetic and artistic value of old maps to the information value.

Based on the information mentioned above was decided to apply process of digital cataloguing and digitization of map archive of Institute of Geography and also making them accessible. Especially easy accessibility enable usage of old maps by public and also by wide range of experts from various fields of science e.g. landscape planning, climatologist, geographers etc. Cataloguing and digitization in processed within the grant of Ministry of Culture, in cooperation of Moravian Library and Library of Charles University.

2.1 Cataloguing and digitization

Cataloguing and digitization of map sheets consists of few steps based on experience of other institutions and existing literature. E.g. Lee [10] defined basic steps in digitization process as follow:

- Assessment and Selection of Source Material
- Digitization Assessment
- Benchmarking
- Full Digitization
- Quality Assessment
- Post-Editing
- Application of Metadata
- Delivery

2.2 Assessment and Selection of Source Material

Selection of material for digitization is based on two basic criterions. First criterion is constituted by time aspect: 1. Maps are digitized from the oldest to the newer. Second used criterion is necessity of existence of digital copy: 2. Priority digitization is used for newer in case if there is a need for digital copy.

2.3 Digitalization

Process of digitization itself is based on sheet-fed scanner Contex Chameleon G 600. Technical specification of scanner: Pixels: 77,000, max resolution: 9600 (in dpi), Scan Accuracy: 0.1%+1pixel, Scanning Speed inch/sec: (400 dpi turbo, 24-bit RGB) 0.6”/s 1.0”/s, Max Scan Width: 36” (914 mm), Max Media Width: 44” (1118 mm), unlimited scan length.

Resolution used for scanning is four hundreds dots per inch with twenty-four bit colour depth. This was determined as adequate for reproduction of facsimiles as identical copies (Brůna [1]). Digitized maps are saved on the data arrays as tiff image files with lossless compression.

Every map sheet is during scanning process manually controlled due to possibility of existence of errors.

2.4 Metadata

Every map sheet contains added metadata information compatible with metadata standard used in national library of the Czech Republic and adopted by libraries of the Masaryk University.

Minimal record of MARC 21 for Union Catalogue of the Czech Republic (cartographic document) [12]:

001 Control Number
003 Identifier of Control Number
005 Date of last record
008 Fixed length data fields
020 International standard number (ISBN)
040 Source of cataloguing
041 Code of language
044 Code of country
072 Code of subject category
080 International categorization
100 Main Heading – personal name
110 Main Heading – name of corporation
111 Main Heading – name of action
245 Title information
250 Publishing information
255 Mathematical information
260 Publisher information
300 Physical decryption
490 Edition information
505 Formalized note to content
700 Subheading – personal name
710 Subheading - name of corporation
711 Subheading - name of action
910 Information for Union Catalogue

2.5 Accessibility

Digitized materials of mentioned map archive are still not easily accessible. Currently has to be decided the way of accessibility. There are many ways haw to make digitized materials accessible. It is presupposed that there will be two main level of access. First level will be represented by free access via internet (lower resolution, watermarking etc.). Second level will be represented by possibility of downloading of copies in original resolution (FTP or send by mail on DVD etc.).

2.6 Future

Future brings question about several issues. One of them is Long term preservation. It is clear, that frequent changing of used formats, software, hardware and also employees can be considered as threads to the readability of digitized materials. Several strategies of long-term digital preservation are suggested in Krebs and Borghoff [9].

3. OLD MAP OBJECTS THESAURUS

The aim of Old Map Object Thesaurus (OMOT) is to provide the list of object types which used to be displayed in old maps together with their definitions and hierarchical relations. For example, such thesaurus will certainly contain high-level terms like water body, forest, human settlement, road, relief, but also low-level terms like swamp, vineyard, castle, or mountain. It is obvious that these object types can be considered also as map layers from contemporary point of view.

Developing OMOT must be based on long term study of large variety of old maps. Map archive of Institute of Geography, Masaryk University, with 18000 maps seems to be perfect candidate for this purpose. Together with studying old maps content, it seems advisable to take advantage of existing thesauri or spatial classifications related to spatial objects. At the moment, we focus on three of them: CORINE Land Cover classification, FAO Land Cover Classification System, and GEMET thesaurus.

CORINE Land Cover (CLC) classification was designed by European Environment Agency within project Coordination of Information on the Environment (CORINE) [4]. Although it was designed for classifying land cover from satellite imagery, it is possible to use CLC for general land cover classification. CLC has strict three-level hierarchical structure with 5 items in the highest level and 44 items in the lowest level. However, CLC was designed for output scale 1:100 000 with the smallest mapping unit 25 hectares (and at least 100m width of linear features), so it can not contain all potential OMOT types such as roads or buildings.

Land Cover Classification System (LCCS) is a classification system developed by Food and Agriculture Organization of United Nations (FAO) [2]. Unlike CLC classification, LCCS is complete specification of the land cover classification process. The process consists of two phases. In dichotomous phase, eight major land cover types are distinguished. In modular-hierarchical phase, the set of classifiers is tailored to the major land cover type. This approach reduces the total number of impractical combinations of classifiers.

GEMET (General Multilingual Environmental Thesaurus) thesaurus was designed by European Topic Centre on Catalogue of Data Sources and European Environmental Agency as a compilation of the best of the presently available excellent multilingual thesauri related to environment [5]. GEMET contains around 5000 terms translated into 27 languages and arranged in 4 super-groups, 32 groups and 40 themes, so it is much more complex than CLC. Furthermore, GEMET does not contain only spatial objects, but also general terms related to environment, so not all terms are suitable for OMOT.

4. OLS MAP INDICES THESAURUS

For cataloguing and preserving purposes it seems reasonable to create database of old maps and map series, so called Old Map Indices Thesaurus (OMIT). Such database is important at least for two reasons. First, imagine a situation when a librarian is cataloguing some old map which was previously catalogued in another library. Using OMIT, the librarian could just find existing

data about the map and use it instead of building it from the ground. Second, imagine a situation when a librarian digitizes some old maps. If something like OMIT exists, the library could just find if another library does not have the same map already digitized.

The OMIT should contain information similar to bibliographic record with special attention paid to the title, author, spatial extent, spatial reference system, etc. The OMIT must be also prepared to store information about map series including set of map sheets, their codes, spatial extension, etc.

5. CONCLUSIONS AND FUTURE WORKS

Presented ideas of Old Map Objects and Old map Indices Thesauri can improve searching for old maps itself, old maps objects and also increase usability of old maps information.

Each of the existing classifications and thesauri, CLC, LCCS, and GEMET, can be beneficial for OMOT. The CLC seems to be good starting point of high level and middle level terms structure. The LCCS may be used in similar way, but its second modular-hierarchical phase is potentially very useful for low-level terms. Finally, the GEMET thesaurus contains general environmental feature types, not only land cover types. It seems reasonable to profit from GEMET multilingualism too. It will be necessary to make supplement and structure adjustment to these existing classifications to cover most of objects which can possibly exist on old maps.

The first step to the OMIT creation is a deep study of old maps and old map catalogues followed by design of a conceptual data model. The very important step is also to verify the model by transforming existing map databases into the model, i.e. by harmonisation. The OMIT will be designed in cooperation with Moravian Library, Brno.

This research is supported by funding from the Ministry of Culture, the Czech Republic.

LITERATURE

[1] Brůna V., 2002. Identifikace historické sítě prvků ekologické stability krajiny na mapách vojenských mapování. Acta Universitatis Purkynae. Studia Geoinformatica II. Ústí nad Labem. (In Czech)

[2] Di Gregorio, A., Jansen, L.J.M., 2000. Land Cover Classification System (LCCS): Classification concepts and user manual. 179p. FAO, Rome.

[3] Dublin Core Metadata Initiative, 2011. DCMI Specifications. Available at: http://dublincore.org/specifications/ [accessed 15 February 2011].

[4] European Envorinmental Agency, 1994. CORINE Land Cover. Commission of the European Communities. 163p.

[5] European Environment Agency, 2011. GEneral Multilingual Environmental Thesaurus. Available at: http://www.eionet.europa.eu/gemet [accessed 15 February 2011].

[6] International Organization for Standardization, 2003. ISO 19115 Geographic information - Metadata.

[7] International Organization for Standardization, 2007. ISO 19139 Geographic information - Metadata - XML schema implementation.

[8] Jobst M, Gartner G., 2010. Structural Aspects for the Digital Cartographic Heritage. In Preservation in Digital Cartography, Archiving Aspects, ed. Markus Jobst, Springer, Berlin, ISBN 978-3-642-12732-8.

[9] Krebs N, Boorghoff U.M., 2010. State of the Art Survey of Long-term Archiving Strategies in the Context of Geo-Data/Cartographic Heritage. In Preservation in Digital Cartography, Archiving Aspects, ed. Markus Jobst, Springer, Berlin.

[10] Lee S.,D., 1999. Scoping the Future of the University of Oxford's Digital Library Collections, Final Report, Available at: http://www.bodley.ox.ac.uk/scoping/report.html [Accessed 21 January 2011].

[11] Library of Congress, 2011. MARC Documentation. Available at: http://www.loc.gov/marc/marcdocz.html [accessed 15 February 2011].

[12] National Library of the Czech Republic. Minimal record of MARC 21 for Union Catalogue of the Czech Republic, Available at: http://www.nkp.cz/ [Accessed 11 January 2011].

[13] Open Geospatial Consortium, 2007. OpenGIS Catalogue Services Specification. Available at: http://www.opengeospatial.org/standards/cat [accessed 15 February 2011].