

AUTOMATION OF GEOSPATIAL RASTER DATA ANALYSIS AND METADATA UPDATING: AN IN-DATABASE APPROACH

LUKÁŠ BRŮHA

Charles University in Prague, Faculty of Science, Department of Applied Geoinformatics and Cartography, Czech Republic

ABSTRACT

This paper proposes a spatial data infrastructure (SDI) module for management of a continuous flow of geospatial images and related metadata. Examples of such flows are continuously acquired map scans from the digitalization process of an old maps collection, or the satellite imagery retrieved through a receiving station. Storage of the raster data in a database is a key feature of the system, which enhances the usual tasks and usability of SDI systems. The analytical procedures deployed within the data store perform automated raster analysis and content-based metadata extraction. This functionality is illustrated with two experiments – improving the display of early map scans and snow and cloud detection from satellite images. Applications of the proposed approach and utilization of the prototype application by geographers and cartographers are discussed.

Keywords: SDI, image data, metadata, in-database, automation

1. Introduction

At the Faculty of Science of Charles University in Prague, huge amount of descriptive, statistical and geometric spatial data is used for research and education purposes. A Spatial Data Infrastructure (SDI) implements a complex framework of technology, geographic data, metadata and users in order to use spatial data in an efficient way. Recently, the amount of spatial raster data has grown significantly due to a new receiving station of satellite imagery and the advancement of the old maps collection digitalization. The increased pressure on human and technological resources unveiled, how far the current approaches and tools designed for vector data are unsuitable for raster images, that are much bigger in data volumes and more variable in storage formats. This implicates higher demands on management and administration of the data store mechanism.

1.1 Related work

Data and metadata records are two crucial components in any SDI. Metadata enable data discovery and access for users and provide information about the purpose, currency and accuracy of spatial data sets (Olfat 2013). However, the manual creation, update and authoring of images' metadata is considered as being monotonous, time-consuming, and labor-intensive tasks (Trilles 2012). That is why challenges arise regarding metadata collection, storage, updating and integration in metadata catalogues (Batcheller 2009; Grill 2009; Olfat 2013).

Therefore, there is a need for an automated administration of collected raster data and related metadata that is more efficient than current archiving approaches. The

key idea to increase the effectiveness of existing SDI solutions is the application of analytical tools for raster data integrated within the archiving system. This shift of application logic is allowed due to newly introduced support to in-database storage of raster data by several database management system (DBMS) vendors (PostgreSQL 2013; Oracle 2013).

The effectiveness of this approach is multiplied by having all available data in one place within SDI, enabling the extrusion of regions in the images or intersection analysis using available vector data. It also performs the vital functions that make spatial data interoperable, i.e., capable of being shared between systems.

To increase the effectiveness of searching for the desired raster data, the content-based metadata of such an image are needed. Their creation and retrieval fills the gap between low-level information that can be processed by computers and high-level semantic information understandable and applicable by humans (Akrivas 2007; Zhang 2012). The application of analytical tools in a raster processing line can automate the generation of such metadata or image annotation.

To prove the functionality of the proposed solution and to demonstrate the potential usage two case studies are presented in this paper.

The **Floreo** (Demonstration of ESA Environments in support to FLOod Risk Earth Observation monitoring) research project and the existence of the receiving station for satellite data at the Faculty of Science were the motivation for an implementation of clouds and snow detection procedure. The receiving station continuously provides AVHRR/NOAA images. Variety of snow and clouds detection methods using the Advanced Very High Resolution Radiometer (AVHRR) data have been already

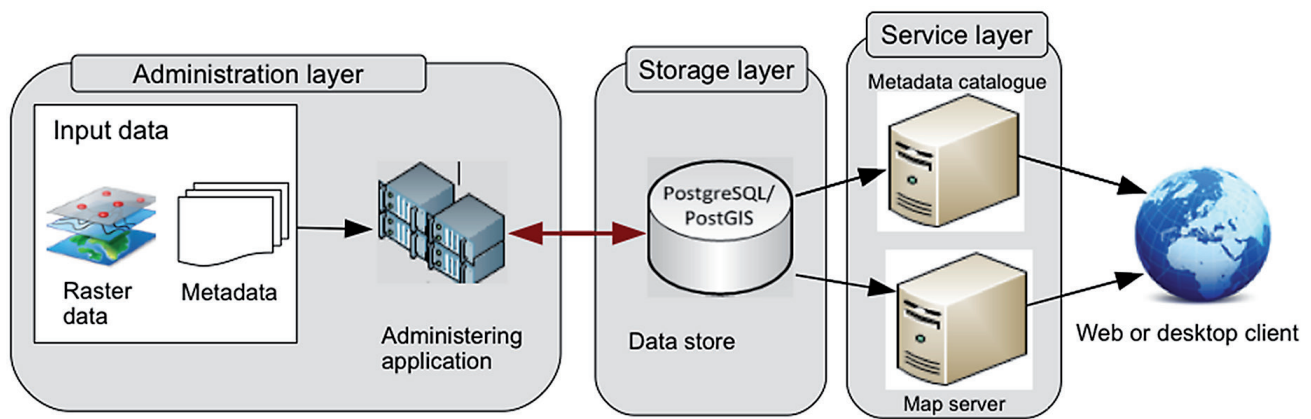


Fig. 1 Implementation architecture for automated raster data management.

reported. While it is relatively uncomplicated to separate snow-free land from snow-covered land using spectral characteristics, it is no easy task to discriminate between snow and clouds (Höppner 2002). The variety of works referred to extraction of snow or clouds from NOAA/AVHRR (Allen 1990; Gesell 1989; Saunders 1986; Simpson 1998; Voigt 1999).

The **TEMAP** (Technology for discovering of map collections) research project (TEMAP 2014) aims at applying the advancements in geospatial web technologies to facilitate the access to early maps for the end-user. The map collection of the Faculty of Science, Charles University in Prague contains tens of thousands of maps with more than 35,000 already digitized and catalogued. As an example of similar initiatives, the David Rumsey Map Collection (2014), can be mentioned. It contains more than 150,000 maps, of which 42,000 are digitized and georeferenced.

Solutions being developed within TEMAP project adapt and further extend the latest technologies for searching and distribution of digitized maps like MapRank (2013), enabling geographic searching by map location and coverage in Google Maps, or Georeferencer, designed for crowdsourced georeferencing for map collections (Fleet 2012).

1.2 SDI module for raster data management – implementation architecture

The initial work on SDI solution was introduced by Hettler (2012) to provide means for automatic management of continuously acquired raster data and metadata. The implementation architecture consists of several components, as depicted in Figure 1.

The administration layer provides an environment for the initialization and configuration of the solution for automatic raster data and metadata archiving and publishing. Technically, the administration layer is based on the Java application *MtdtRasPub*, which, in addition to controlling the raster data flow, constitutes the metadata record for each raster image from all available sources

(World Files, raster headers, bibliographic records). Metadata records follow the ISO 19115:2003 standard, the current “best practice” standard defining the geospatial metadata format.

The storage layer, which includes the databases used to store data and metadata, is based on the PostgreSQL database platform. Within this layer, appropriate data structures for data and metadata are built in order to perform their automatic publication for the system’s users.

The service layer manages the communication of the data store with the metadata catalogue and map server, employing for this sake the GeoNetwork opensource (GeoNetwork 2013) metadata catalogue and GeoServer (Geoserver 2013) map server. The web-based graphic user interface presents the data and metadata to user.

This SDI module for raster data administration provided the automation of raster data storage and distribution via the web user interface together with all available metadata. However, using such a solution for analytical image processing, the data must be first transferred from the data store to an external application. Thus, so as to fully exploit the advantage of storage of raster data in the database, an extension of this architecture is needed. Consequently, the next section presents the shift of image processing from an external application onto the storage layer and discusses the requirements for the deployment of an image processing functionality within the DBMS. Furthermore, case studies of a custom-made images’ analysis functionality are presented, followed by a discussion on the potential usage of a solution for a broad geographic community.

2. In-database image analysis approach

The current usual practice is based on the out-of-the-database storage of raster data. In a relational database only the metadata describing the image are kept. For any processing or analysis, the raster data must be first transferred to separate processing and analytical software applications, as it is suggested in Figure 2. The analytical

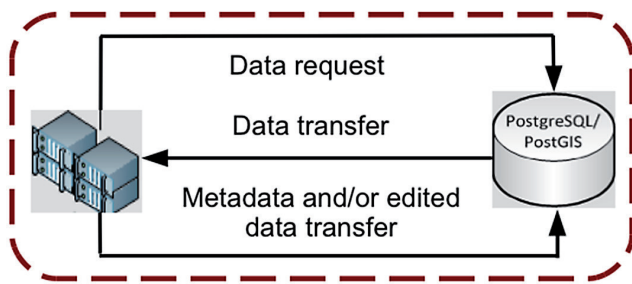


Fig. 2 Raster analysis in detached data storage and data analysis applications.

task can result in raster editing. In such a case, a new raster representation needs to be transferred back to the data store, causing extra data transfer overhead.

The advantage of keeping the data out-of-the-database, i.e., as *binary large objects*, appears in case the publication of stored raster data from the database system (like map service publishing) is the only objective of an application. The first performance evaluation results presented in (Hettler 2012) shows that the publication from a native PostGIS raster format is slightly slower than from the alternative binary raster storage, which usage however prevents from the employment of analytical tools.

In-database storage. The in-database strategy (Xie 2013; Obe 2011) employed by the solution for geospatial images proposed in this paper has several features to enhance the storage and analysis of big geospatial images.

The first feature is moving the image processing closer to the data to avoid moving large data sets from the databases to detached analytical software. The second feature is parallel processing provided by the database for the in-database raster format storage. The third feature is concurrent processing that enables leveraging the power of computer clusters to concurrently process numerous images (Xie 2013). In the Figure 3, the retrieval of results of an image analysis, which is performed on the database side, is depicted.

In-database image processing functions. There are countless of possible image analysis functions. Raster data processing and analysis involves a large set of operations, such as radiometric and geometric corrections, image transformation and mosaicking, image enhancement,

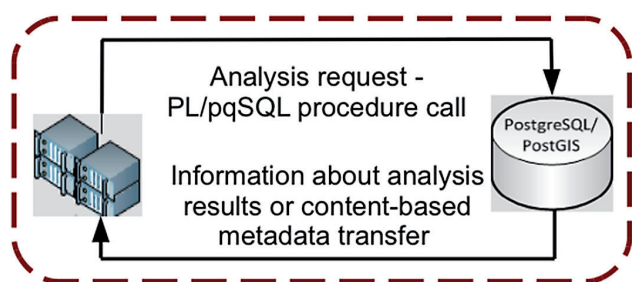


Fig. 3 The character of raster data stored in-database processing.

pattern recognition and raster map algebra, to name a few (Gonzales 2006).

Database platforms with raster data support implement only core functions that are required by database management or that improve the effectiveness of data manipulation, such as image updates, processing or aggregation. These complement traditional GIS applications and can be reused by complex or custom-made analytical procedures deployed for a specific purpose, providing effective raster data manipulation for such developed procedures.

In order to fully utilize the effectiveness of in-database approach for analytical procedures over spatial raster data, the database platform is supposed to support the following key functionality:

- raster bands accessors,
- raster pixel accessors and setters,
- raster band statistics,
- map algebra over individual pixels,
- spatial indexing,
- datum definition and coordinate system transformation.

This allows for a basic analysis and moreover supports the development of custom-made functionality, providing procedural language. With respect to the architecture presented above, PostgreSQL with spatial extension PostGIS (PostGIS 2013) was chosen for the implementation of the proposed SDI enhancement. PostgreSQL offers this key functionality for raster data processing and also provides PL/pgSQL procedural language.

3. Experiment

Ongoing research projects like TEMAP, which aims to develop of technologies and procedures for discovering old maps collections, or FLOREO, which is concerned with snow detection from satellite images, provided the motivation and data for the tests of the proposed solution. Continuous flow of acquired very large raster data from the satellite images receiving station and from the digitalization of tens of thousands old maps required the greatest possible reduction of raster data movements for the sake of processing and analysis.

This requirement was met by employing the in-database approach and the development of specialized analytical functions within the data store. Implementation of this functionality is enabled through the procedural languages. The PL/pgSQL procedural language of the PostgreSQL database platform was utilized for these purposes.

3.1 Cloud and snow detection

The algorithm for snow detection introduced within the FLOREO project was designed based on the past work of Romanov (2000). The cloud detection part is adopted from the AVHRR Processing over Land Cloud

and Ocean (APOLLO), which was developed by Saunders (1986). The implementation of such procedures within the data store aimed at a retrieval of basic information about the snow and cloud coverage in the image. This information is utilized for two purposes. First, the automatic identification of such images appropriate for classification. Second, the automatic creation of content-based metadata to increase the effectiveness of the search in the metadata catalogue.

NOAA-AVHRR data. NOAA is a polar satellite that circles at an altitude of approximately 850 km. The satellite scans each place on Earth at least twice a day, with increasing frequency at places closer to poles. The Advanced Very High Resolution Radiometer (AVHRR)

instrument on board of NOAA has 5 (or 6) wavelength channels. The channels are optimized to measure cloud and surface characteristics with minimum contamination from other atmospheric constituents. The channel specifications are presented in Table 1.

Tab. 1 AVHRR/3 channel characteristics.

Channel number	Resolution at Nadir	Wavelength (μm)
1	1.09 km	0.58–0.68
2	1.09 km	0.725–1.00
3a	1.09 km	1.58–1.64
3b	1.09 km	3.55–3.93
4	1.09 km	10.30–11.30
5	1.09 km	11.50–12.50

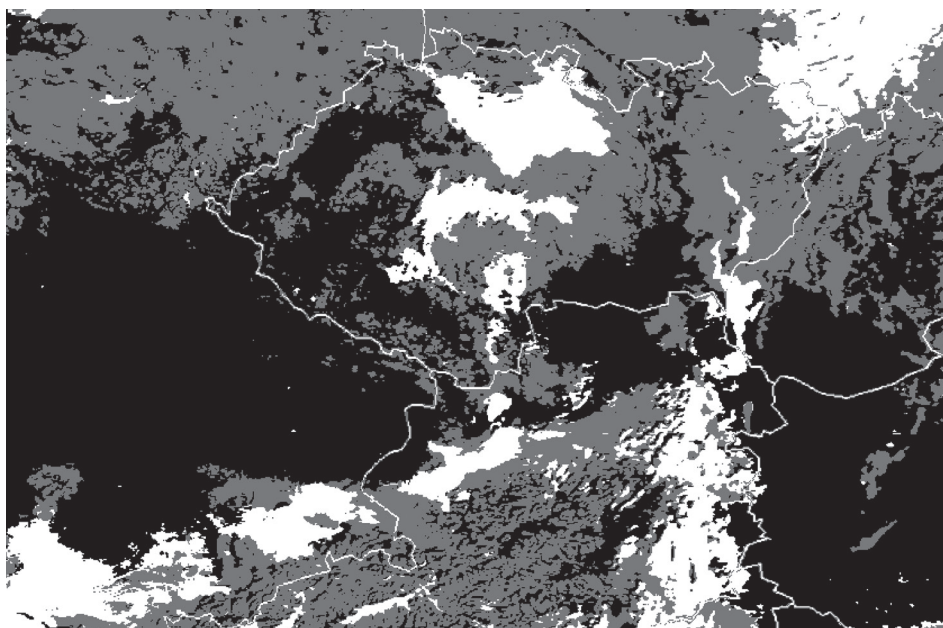
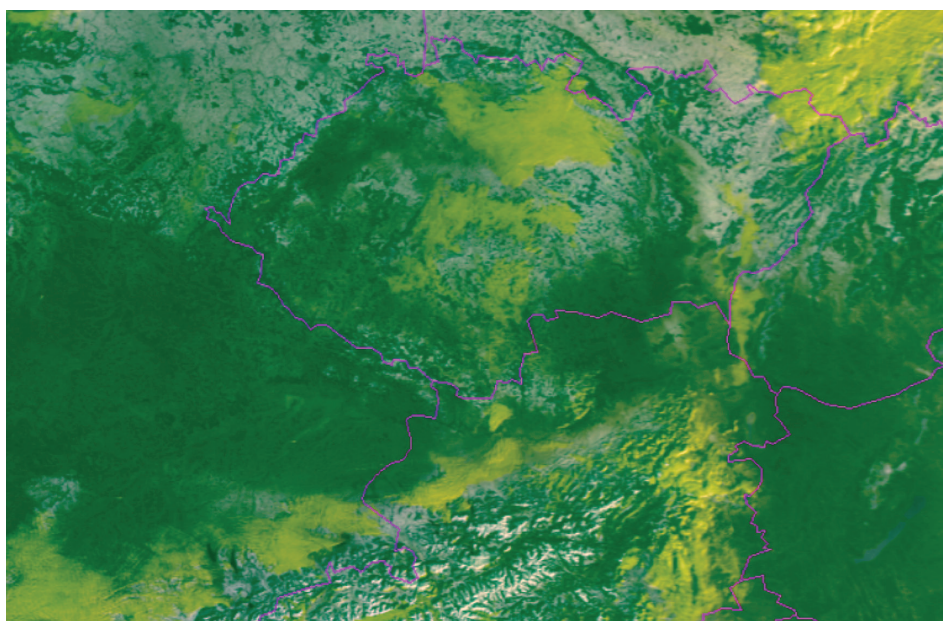


Fig. 4 The classification of the original NOAA image (a) into the categories of land, snow and cloud coverage (b).

The spectral signatures of snow and clouds can be very similar and depend on various environmental factors. The snow and cloud discrimination relies on threshold value estimates, like minimum possible surface temperature of snow compared to clouds, as determined by a histogram analysis of temperature and reflectivity. Threshold values are instrument specific.

Snow detection. The snow detection procedure is implemented based on a series of tests adopted from the algorithm developed by Romanov (2000). An image pixel is identified as snow by a threshold method, which tests if the signal in a channel or combination of channels corresponds to defined spectral characteristics in a cloud free atmosphere. The main tests for daytime are:

1. $NDSI > NDSIT$
2. $SI > SIT$

where SIT and $NDSIT$ are threshold values, $SI = R0.6/R1.6$ is a snow index, $NDSI = (R0.6 - R1.6)/(R0.6 + R1.6)$ is a normalized difference snow index and $R\lambda$ is the reflectivity in the spectral channel denoted by λ .

Cloud detection. The PL/pgSQL procedure for cloud detection follows the threshold-based approach of (Saunders 1986). The procedure tests the signal against the thresholds. The scene is marked as cloud contaminated if one of the following conditions for daytime is met (Roebeling 2003):

1. $T10.8$ (measured) $< T10.8$ (cloud free) – temperature threshold (Temperature test)
2. $R0.6 > R0.6$ (cloud free) + reflectivity threshold (Reflectivity test)
3. $T10.8 - T11.9 > T10.8$ (cloud free) – $T11.9$ (cloud free) + $S-T$ (Semi-transparency test)

$S-T$ is the semi-transparency threshold, $T\lambda$ denotes the equivalent of black body temperature in the spectral channel λ .

The original NOAA image, see Figure 4(a), classified into three categories – land, snow and cloud coverage, can be seen in Figure 4 (b). The content-based metadata describing the percentage of individual categories is the outcome from procedures, subsequently encoded within the ISO 19115:2003 metadata record.

3.2 Improving raster display

In image processing, the normalization is an image enhancement technique that improves the contrast in an image by stretching the range of pixel intensity values into a desired range of values (Gonzales 2006).

Implementation of a procedure for an improvement of raster display is motivated by the effort to increase the readability of early maps scans. The original documents faded out due to the long period of time since their creation, the character of the materials used and colouring techniques.

The PL/pgSQL procedure transforms an image I with intensity values in the range (Min, Max) , into a new image

IN with intensity values in the range $(newMin, newMax)$. The linear normalization is performed according to the formula:

$$IN = (I - Min)(newMax - newMin)(Max - Min) + newMin.$$

The Figure 5 illustrates the outcome of the histogram stretching function. The implementation of such a procedure is straightforward and also computationally efficient. That is due to the optimized functions for raster data manipulation, editing and yielding of image statistics or image histogram. These functions are natively provided by the PostgreSQL Raster platform and are re-usable within other developed functions.

3.3 Metadata

GeoNetwork opensource metadata catalogue is employed by the proposed solution to provide means of description of various types of geographic data (vector or raster layers, map services, statistical data).

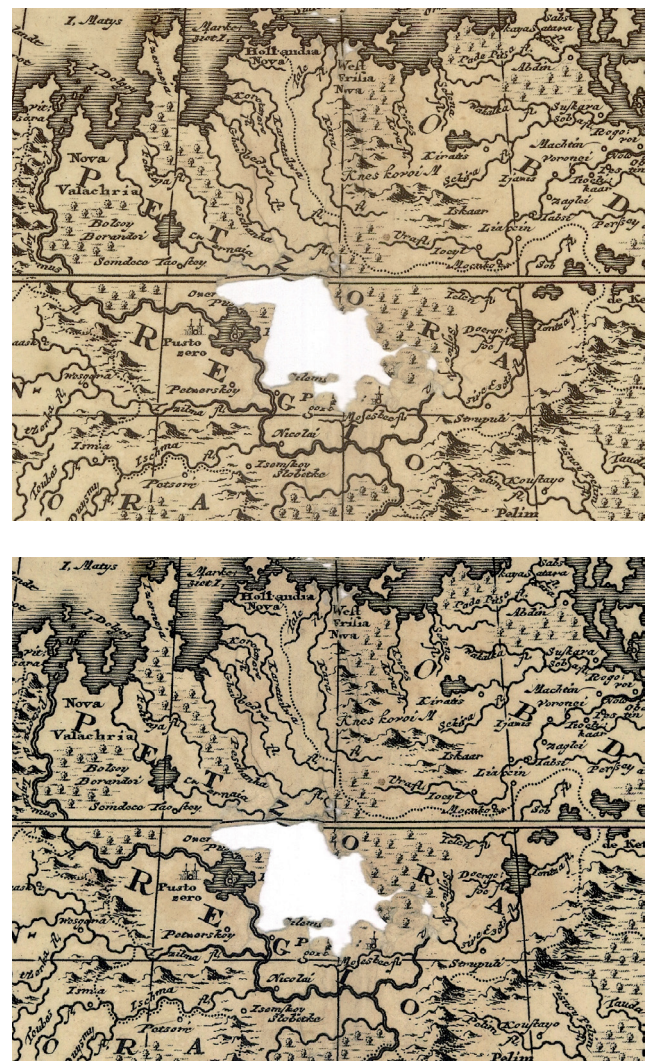


Fig. 5 The segment of an early map with some damages, changes of paper colouring and faded labelling. (a) Before and (b) after application of normalization.

The metadata document is formed within the administration application from available sources in accordance with ISO 19115:2003 rules. Only a small portion of elements is used following the GeoNetwork and INSPIRE (2014) recommendations on required or highly recommended elements to properly describe geographic data. The compliance of metadata with these rules is checked by GeoNetwork when metadata records are imported or updated. For this sake, the GeoNetwork's `xml.metadata.insert` service is employed. Three groups of metadata fields in a resulting metadata record can be identified based on the source of their origin.

Available descriptive metadata. Depending on the data source, the descriptive metadata can be retrieved from World Files, raster headers or bibliographic records. To carry this out automatically, the format of the source document must be known, i.e., a catalogued description of an old map in XML document following the MARC 21 standard. The cataloguing procedure itself follows the methodology described in Novotná (2013) facilitating the identification of the key corresponding fields in both standards – the source MARC 21 bibliographical standard and the target ISO 19115:2003 geo-informatical standard. The following fields are acquired by processing the source documents:

Title, Date, Date type, Abstract, Purpose, Descriptive Keywords, Language, Topic category, Scale Denominator, Temporal Extent, Geographic Bounding Box, Reference System Info.

System generated metadata. Metadata generated automatically by some components of the system like *Online resource linkage* (url of the source disseminated by GeoServer) or *Data quality info* belongs to this category. This category includes metadata fields, whose values are set to the system administration application by a person responsible for the dataset, like *Presentation form, Organization name, Role, Maintenance* and *update frequency*. Also, the values of fields like *Abstract* or *Purpose* can be determined using the objective of the specific satellite mission and further automatically set by the administration MtdtRasPub application for all images of the dataset.

Metadata as an analysis product. In the field *Supplemental Information*, additional information acquired through raster analysis is encoded. This refers to the percentage of cloud and snow coverage in satellite images and also to the original and new minimum and maximum intensity values of old maps scans. The unknown reference system or map scale of old maps can stand as another example. In this case such information cannot be retrieved from bibliographic document. Cartometric analysis of such an old map scan however can provide estimates of these parameters (Bayer 2014).

The assignment of metadata fields into the categories above is not strict and depends on the unique

characteristics of datasets. The update of existing metadata records is enabled by the system and carried out on an 'as needed' basis applying the *metadata.insert* xml service. The unique identifier prevents from the creation of duplicated records. The updating of existing metadata is required to encode the analysis results.

4. Conclusions

An SDI module for management of continuous flow of raster data and related metadata was proposed. This module addresses the needs for the automation of raster data archiving, analysis and distribution.

System evaluation. The functional parts of the prototype system have been developed in cooperation with the researchers and end-users of provided services.

The developers of the metadata solution, along with map archivists and end-users, defined fields from bibliographic records, that would be relevant for both public and scientists working in fields of geography and cartography. Selected metadata elements provided sufficient map description and search capabilities within the SDI system.

Also, the role of content-based image description was proven to be the key for the effective management of satellite imagery. Due to huge data volumes regularly produced, older or unused images are moved to be archived on backup media like magnetic tape. The image description is then crucial in allowing the searching of archived images, whose display is not available on-line.

The analysis of snow and cloud coverage demonstrated the way of content-based metadata creation. Romanov (2000) presented evaluation of classifications for satellite-based snow products. Results of the method varied from 75% to 85% of correct classification depending on environmental conditions. The comparison of results obtained by manual processing with automatically acquired results fit into this range. Nevertheless, to improve the image search capabilities and to fully answer the needs of end-users, more complex analysis on snow and clouds characteristics and more detail placement of such phenomenon are necessary.

The presented experiments provided valuable results for related research projects. The retrieved metadata enhanced the search capabilities and presented information about the suitability of an image for further analyses (like the classification of snow characteristics). The normalization of old map scans improved the readability of such documents. The main contribution, however, inheres in proving the functionality of the in-database analysis approach followed up by the automatic content-based metadata creation which is a major open research problem, not only to metadata catalogue or SDI systems but in other fields too.

Future development. The introduced approach provides many opportunities for geographers of various specializations to facilitate the retrieval of information from huge rasters, share it and effectively search for available data within the SDI.

The extension of the automatic processing of old map scan is an example. With a reasonable amount of effort, analytical procedures for the determination of the level of damage of a historical document, like that in Figure 5, can be implemented.

As another example, the enhancement of a geo-referenced mosaic created by historical cartographers from early map series like the military survey (Molnár 2011), can be mentioned. The image histogram equalization would provide a unified appearance by removing differences in contrast between individual lists caused by different materials or archiving approaches. More advanced procedures may deal with automatic map field extraction for such map series.

Another related example of potential future development is the extension of proposed snow and cloud detection aiming at the automation of snow and cloud typology classification or land use classification within the data store.

As shown in Dang (2012) the in-database storage is also a promising approach for the effective distribution and visualization of data changing continuously in time and space. Examples of such phenomena are temperature, pressure, precipitation, snow cover, land use or population density.

Future steps in application development will aim at the implementation of additional analytical procedures: (a) the image histogram equalization for the sake of mosaicking, (b) the embedding of existing procedures on cartometric analysis into the SDI system and (c) the integration of the application with solutions for crowdsourced georeferencing.

Acknowledgements

This work has been supported by the TEMAP project of the NAKI program with DF11P01OVV003 code.

REFERENCES

- AKRIVAS, G., PAPADOPOULOS, G. T., DOUZE, M., HEINECKE, J., O'CONNOR, N. E., SAATHOFF, C., WADDINGTON, S. (2007): Knowledge-based semantic annotation and retrieval of multimedia content. In: SAMT 2007 – 2nd International Conference on Semantic and Digital Media Technologies, 5–7 December 2007, Genoa, Italy.
- ALLEN, R. C., DURKEE, P. A., WASH, C. H. (1990): Snow/Cloud Discrimination with Multispectral Satellite Measurements. *Journal of Applied Meteorology*, Vol. 29, pp. 994–1004.
- BATCHELLER, J. K., GITTINGS, B. M., DUNFEY, R. I. (2009): A Method for Automating Geospatial Dataset Metadata. *Future Internet* 1, no. 1, 28–46, <http://dx.doi.org/10.3390/fi1010028>.
- BAYER, T. (2014): Estimation of an unknown cartographic projection and its parameters from the map. *Geoinformatica*, March 2014, <http://dx.doi.org/10.1007/s10707-013-0200-4>.
- DANG, T. N. U. (2012): PostGIS Raster: Extending PostgreSQL for The Support of Continuous Fields. Master thesis, Université Libre de Bruxelles, Belgium.
- DAVID RUMSEY MAP COLLECTION. Homepage [online]. Available from: <http://www.davidrumsey.com> [Accessed 12 May 2014].
- FLEET, C., Kimberley C. K., PŘIDAL, P. Georeferencer: Crowdsourced Georeferencing for Map Library Collections. *D-Lib Magazine: The Magazine of Digital Library research* [online]. 2012, Vol. 18, No. 11/12 [cit. 2013-01-03]. <http://dx.doi.org/10.1045/november2012-fleet>.
- GEONETWORK, 2013. Homepage [online]. Available from: <http://geonetwork-opensource.org> [Accessed 18 October 2013].
- GEOSERVER, 2013. Homepage [online]. Available from: <http://docs.geoserver.org/> [Accessed 12 October 2013].
- GESELL, G. (1989): An Algorithm for Snow and Ice Detection Using AVHRR Data – An Extension to the APOLLO Software Package. *International Journal of Remote Sensing*, Vol. 10, No. 4, pp. 897–905, <http://dx.doi.org/10.1080/01431168908903929>.
- GONZALES, R. C., WOODS, R. E. (2006): *Digital Image Processing* (3rd Edition). Prentice-Hall, Inc. Upper Saddle River, NJ, USA.
- GRILL, S., BAYER, T., SCHNEIDER, M., JEDLIČKA, J. (2010): Archive and catalogue system for receiving satellite data as a part of academic SDI. In: Manakos, I., Kalaitzidis, Ch. *Imagin[e,g] Europe*.
- HETTLER, J. (2012): Automatická publikace metadata a dat pro mapové a katalogové systémy z rastrových podkladů v PostgreSQL. Master Thesis, Charles University in Prague, Czech Rep.
- HÖPPNER, E., PRECHTEL, N. (2002): Snow Cover Mapping with NOAA-AVHRR Images in the Scope of an Environmental GIS Project for the Russian Altai (South Siberia). – Proc. of the ICA Comm. on Mountain Cartography, Mt. Hood, Oregon, May 15–19, 2002, http://www.karto.ethz.ch/ica-cmc/mt_hood/proceedings.html, 26 p.
- INSPIRE directive, 2014. Homepage [online]. Available from: <http://inspire.ec.europa.eu/> [Accessed 13. May 2014].
- MAPRANK SEARCH. Available from: <http://www.klokantech.com/mapranksearch> [Accessed 12 October 2013].
- MOLNÁR, G. (2011): Tvorba georeferencované mozaiky souboru historických map s použitím vymezeného lícování. *Novinky Zeměměřičské knihovny* 2/2011.
- NOVOTNÁ, E. (2013): Staré mapy a grafiky v Geografické bibliografii ČR on-line. *Knihovna: knihovnická revue*, 2013, Vol. 24, No. 1, pp. 5–27, <http://full.nkp.cz/nkkr/knihovna131/13105.htm>.
- OBE, R., HSU, L. (2011): PostGIS 2.0 3D and Raster support enhancements. [online]. Available from: http://www.postgis.us/downloads/ncgis2011/NCGISSDBPostGIS20_2011.pdf [Accessed 21 October 2013].
- OLFAT, H., KALANTARI, M., RAJABIFARD, A., SENOT, H., WILLIAMSON, I. P. (2013): A GML-based approach to automate spatial metadata updating, *International Journal of Geographical Information Science*, 27(2), 231–250, <http://dx.doi.org/10.1080/13658816.2012.678853>.
- Oracle, 2013. Oracle GeoRaster homepage [online]. Available from: http://docs.oracle.com/cd/B19306_01/appdev.102/b14254/geor_intro.htm [Accessed 21 October 2013].

- PostGIS, 2013. Homepage [online]. Available from: <http://postgis.refractor.net/> [Accessed 21 October 2013].
- PostgreSQL, 2013. Homepage [online]. Available from: <http://www.postgresql.org/about/> [Accessed 21 October 2013].
- XIE, Q., CHEN, F., ZHANG, Z., LUCENA, I. (2013): In-database image processing in Oracle Spatial GeoRaster. ASPRS 2013 Annual Conference Baltimore, Maryland.
- ROMANOV, P., GUTMAN, G., CSISZAR, I. (2000): Automated monitoring of snow cover over North America with multispectral satellite data, *Journal of Applied Meteorology*, 39, 1866–1880, [http://dx.doi.org/10.1175/1520-0450\(2000\)039<1866:AMOSCO>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(2000)039<1866:AMOSCO>2.0.CO;2).
- ROEBELING, R. A., FEIJT, A. J., DLHOPOLSKY, R., ROOZEKRANS, H. (2003): MSG cloud products Final report.
- SAUNDERS, R. W. (1986): An automated scheme for the removal of cloud contamination from AVHRR radiances over western Europe. *Int. J. Remote Sens.*, 7, 867–886, <http://dx.doi.org/10.1080/01431168608948896>.
- SIMPSON, J. J., STITT, J. R., SIENKO, M. (1998): Improved Estimates of the Areal Extent of Snow Cover from AVHRR Data. *Journal of Hydrology* 204, pp. 1–23, [http://dx.doi.org/10.1016/S0022-1694\(97\)00087-5](http://dx.doi.org/10.1016/S0022-1694(97)00087-5).
- TEMAP project. Homepage [online]. Available from: <http://www.temap.cz> [Accessed 12. May 2014].
- TRILLES, S., DIAZ, L., GIL, J., HUERTA, J. (2012): Assisted generation and publication of geospatial metadata. *Proc. 15th AGILE Int. Conf. Geographic Information Science: Bridging the Geographic Information Sciences*, 2012.
- VOIGT, S., KOCH, M., BAUMGARTNER, M. F. (1999): A Multichannel Threshold Technique for NOAA AVHRR Data to Monitor the Extent of Snow Cover in the Swiss Alps. *Interactions Between the Cryosphere and Greenhouse*. *Proceedings of IUGG 99 Symposium HS2*, Birmingham, IAHS Publications, No. 256, pp. 35–43.
- ZHANG, D., ISLAM, M., LU, G. (2012): A review on automatic image annotation techniques. <http://dx.doi.org/10.1016/j.patcog.2011.05.013>.

Lukáš Brůha

*Department of Applied Geoinformatics and Cartography
Charles University in Prague, Faculty of Science*

Albertov 6

128 43 Praha 2

Czech Republic

E-mail: bruha1@natur.cuni.cz

RESUMÉ

Automatizace analýzy rastrových geodat a aktualizace metadat v infrastrukturách prostorových dat

Nárůst objemu prostorových dat vytvářených a aktualizovaných v krátkých časových intervalech vyžaduje nové přístupy ke správě takových dat. Příkladem může být průběžné topografické mapování, pořizování družicových či leteckých snímků, či digitalizace archivních dokumentů obsahující prostorová data, jako jsou staré mapy.

Článek proto prezentuje modul infrastruktury prostorových dat (SDI) pro řízení datového toku prostorově referencovaných snímků a jejich metadat. Je představen metadatový model, který je automaticky vytvářen z dostupných zdrojů metadat jako hlavičky rastrových souborů, pomocné soubory nebo bibliografické záznamy. Metadatové záznamy jsou klíčovým nástrojem pro vyhledávání dat. Dostupné zdroje metadat však často postrádají dostatečnou charakteristiku obsahu snímku, jehož automatizovaný popis je úzce svázán s analýzou dat. Navržený systém proto integruje archivní a distribuční funkce s funkcemi analytickými a současně poskytuje významnou datovou interoperabilitu tak, aby umožnil práci s různými datovými typy a standardy.

Vzhledem k obrovským objemům zejména rastrových dat je též nutné navrhnout systém tak, aby byl minimalizován transfer dat. Použitý způsob uložení rastrových dat s využitím nativního formátu databáze využívá data uložená v rámci datového zdroje a analytické procedury implementované také v rámci datového zdroje. Tím se minimalizují transfery takovýchto dat z úložiště k analytické aplikaci.

Popsané řešení je demonstrováno na dvou případových studiích. První je získávání informace o podílu oblačnosti a sněhové pokrývky v družicových snímcích NOAA-AVHRR. Druhou úprava kontrastu ve skenech starých map pocházejících z mapové sbírky Přírodovědecké fakulty Univerzity Karlovy. V závěru jsou diskutovány možnosti užití navrženého řešení v široké geografické a kartografické komunitě.