

# Možnosti uchovávání časových sérií rastrových dat a jejich metadat v nich

Pavel Ševčík

Institut geoinformatiky  
VŠB - Technická univerzita Ostrava  
tř. 17. Listopadu 15  
708 33 Ostrava - Poruba  
E - mail: [pav.sevcik@post.cz](mailto:pav.sevcik@post.cz)

## ABSTRAKT

Tato bakalářská práce se zabývá možnostmi archivace časových sérií rastrových dat, včetně jejich metadat. Jedná se o data z různých, především však meteorologických modelů, která jsou periodicky produkována a ukládána do rastrových souborů. Počet a objem dat neustále roste a vzniká tak potřeba jejich efektivní správy. Navržené řešení využívá grafickou knihovnu GDAL pro práci s rastrovými daty a značkovací jazyk XML pro ukládání metadat. Výhodou tohoto řešení je nezávislost na použité platformě a možnost užití v podstatě jakéhokoliv standardu pro metadata ve formátu XML. Pomocí několika navržených skriptů je možno uchovávat, vyhledávat, exportovat a především efektivně spravovat velké množství rastrových dat a jejich metadat.

Klíčová slova : XML, správa dat, rastrová data, geodata

## ABSTRACT

This work deals with the possibilities of archiving the time series of raster data including their metadata. These include mainly (but not only) the data from the meteorological models, which are periodically produced and saved as rasterfiles. Data quantity and data bulk still increase and thus arise the need of their effective administration. Suggested solution uses graphic bookcase GDAL for the work with raster data and markup language XML for storing metadata. The advantage of this solution is dependence on used platform and the possibility of using any standard for metadata in XML format. It is possible through the use of few designed scripts to save, find out, export and mainly effectively administrate a huge amount of raster data and metadata.

Key words: XML, data administration, raster data, geodata

## Úvod

Žijeme v informačním věku a jedna z jeho charakteristických vlastností je neustálý nárůst množství informací. V souvislosti s GIS se můžeme setkat s různými systémy produkujícími velké množství rastrových dat, jejichž obsah se pro dané území periodicky mění. Výsledky modelů a měření se periodicky opakují a ukládají do rastrových souborů. Dalšími zdroji takovýchto dat mohou být například meteorologické radiolokátory, které detekují srážky nad daným územím, nebo družicová měření, jenž jsou schopny v současné době detekovat téměř jakoukoliv látku v atmosféře nebo na zemském povrchu. Tato data jsou následně užívána a analyzována v různých vědních disciplínách jako hydrologie, geologie, meteorologie a dalších. Důsledkem periodického opakování měření je pak velké množství souborů, které je vhodné nějakým způsobem archivovat.

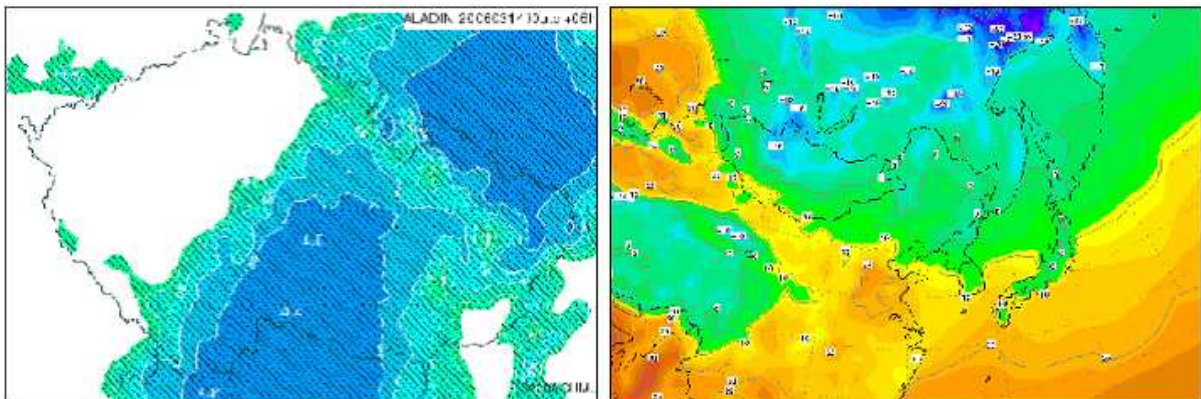
Hlavním tématem této práce je tedy ověření možností archivace takovýchto dat a návrh systému, jenž bude data spravovat a nebude závislý jen na určité platformě a standardu. Základní myšlenkou návrhu řešení je využití grafické knihovny GDAL pro manipulaci s rastrovými daty a značkovacího jazyka XML jako výměnného formátu pro metadata.

Metadata jsou nedílnou součástí samotných dat, popisující jednotlivé vlastnosti datové sady. Nové technologie pronikají i do této oblasti a jako jedno z velmi vhodných řešení se jeví využití výměnného formátu XML s ohledem na nejnovější metadatové standardy. Většina standardů je ovšem časem neustále modifikována nebo přestane být užívána vůbec. Z toho důvodu je vhodné navrhnout řešení problému tak, aby bylo co možná nejméně závislé na použitém standardu. V této práci se využívá standard ISO 19115 a rozšiřující standard ISO/PDTS 19139.4 .

## Úkoly

- Seznámení se s vlastnostmi rastrových formátů používaných v GIS se zaměřením na formáty, které umožňují uchovávat více obrazových kanálů a umožňují zápis metadat pomocí meta-tagů.
- Ověřit možnosti zápisu časové série rastrových souborů do jediného souboru jako jednotlivé kanály s využitím knihovny GDAL nebo jejich utilit. Ověřit možnost zápisu metadat pro jednotlivé kanály.
- S využitím API knihovny GDAL nebo jejich utilit vytvořit skripty umožňující:
  - automatizaci vkládání nových kanálů a metadat do rastrového souboru
  - vyhledávání potřebných rastrových souborů a jednotlivých kanálů
  - export potřebného kanálu do samostatného souboru, popřípadě export pouze výřezu
  - zjištění hodnoty pixlu na zvolené souřadnici,

## Ukázka dat



obrázek č.1 –ukázka výstupů z modelu ALADIN a modelu GFS,  
vpravo: ALADIN - srážky za posledních šest hodin (s vyšrafovanými oblastmi sněžení),  
vlevo : GFS - teplota ve dvou metrech

## Značkovací jazyk XML

Součástí této práce je uchovávaní metadat o rastrových souborech. Bylo tudíž nutné vybrat nejvhodnější způsob ukládání metadat. Jako nejlepší se ukázalo využití jazyka XML. Jazyk XML vytváří jednoduchý, ale velice účinný mechanismus pro ukládání, šíření a zpracování informací. Bylo vyvinut a je podporován konsorciem W3C. Jazyk odděluje data od procesů, má možnost definovat vlastní DTD i vlastní značky v dokumentu. Syntaxe jazyka je poměrně přísná. Jeho hlavní výhodou je, že kromě samotného textu vypovídá i o jeho významu. Dalšími výhodami jsou pak možnost poměrně jednoduché konverze do dalších formátů a nezávislost na používané platformě.

## Možnosti ukládání rastrových souborů

Rastrové soubory se mohou ukládat do relačních databází, které umožňují ukládání rastrových souborů do tzv. geodatabáze. Ukládání do těchto databází je velmi vhodné. Tento způsob ukládání nabízí například databáze ARCSDE od firmy ESRI nebo Oracle 9i. Nicméně pořizovací náklady na tyto produkty jsou velmi vysoké. Nyní se pracuje na tom, aby bylo možné ukládat rastry do databáze PostgreSQL/Postfix. Tato možnost je stále ve vývoji a využívá řadiče PGCHIP grafickou knihovnu GDAL.

Další možností je ukládání těchto dat do souborového systému ve vhodném formátu. Touto možností jsem se zabýval já ve své práci, nicméně i toto řešení sebou nese řadu problémů, například velké množství souborů, neorganizovanost souborů atd. Pro představu je to například 1440 souborů za jeden měsíc pro systém, který každou půlhodinu generuje jeden rastrový soubor. Bylo proto nezbytné navrhnout řešení výše zmíněných problémů a to především pomocí výběru vhodného grafického formátu.

## Volba vhodného grafického formátu

Grafický formát by měl splňovat několik následujících požadavků.

- neproprietární, obecný formát (formát, který není výhradně svázan s nějakým programovým produktem, který jej omezuje)
- možnost zápisu více rastrových souborů do jednoho souboru jako samostatné kanály
- podpora bezztrátové komprese obrazu
- zapisování metadat ve formě metatagů a zápis informací o souřadnicovém systému

Studováno bylo mnoho grafických formátů například Erdas Imagine, GIF i PNG, BIL, BIP, BSQ. Každý z těchto formátů má však určitá omezení. Například u rastrových souborů typu GIF je možno používat pouze 8 bitové palety. U souborů typu PNG je možné ukládat pouze 4 kanály do jednoho souboru. Navíc ani jeden z těchto dvou formátů neumožňuje přímý zápis informací o referenčním systému a vkládání metatagů.

Všechny výše stanovené podmínky nejlépe splňuje pouze jeden formát a to formát TIFF, který je také standardním a výchozím formátem knihovny GDAL.

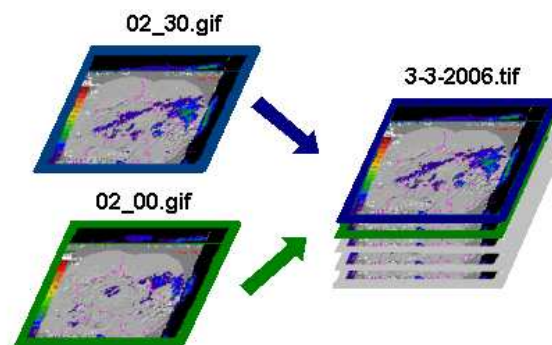
## Grafický formát GeoTIFF

Grafický formát GeoTIFF je podporován většinou aplikací v GIS a je schopen věrně reprezentovat obraz. Podporuje plnou paletu barev a vyšší bitové hloubky. Je schopen přirozené reprezentace dat a rozsáhlých rastrových obrazů. Splňuje požadavky na ukládání jakýchkoliv rastrových dat. Umožňuje přímý zápis informací o referenčním systému a ukládání meta- tagů. Pomocí metatagů je možné vkládat do souboru metadata a to především

metadata týkající se platnosti modelu. Nicméně má i jisté nevýhody. Tou hlavní je velký objem souborů a proto se zavádí různé typy komprese. Použitá komprese je závislá na použité aplikaci. Mezi nejznámější patří Packbits komprese, Huffmanova komprese, LZW komprese a další.

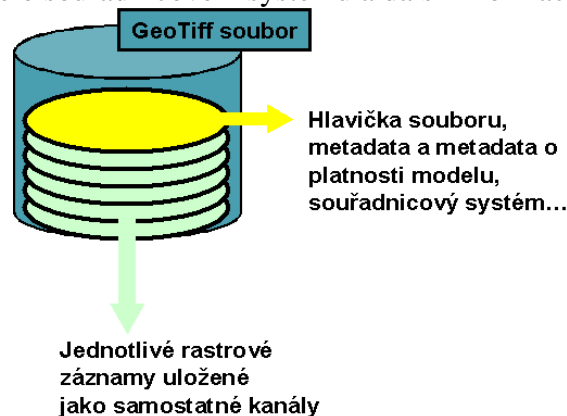
## Návrh způsobu ukládání

Princip skládání rastrových souborů zachycuje následující obrázek.. Zeleně označený soubor byl generován ve dvě hodiny a modře označený soubor o půl hodiny později. V tomto případě se soubory složí do jednoho souboru, jenž obsahuje všechna data za celý den.



obrázek č.2 – princip skládání rastrových souborů

Komprimovaný soubor GeoTIFF obsahuje jednak jednotlivé rastrové záznamy, uložené jako samostatné kanály, dále hlavičku souboru, metadata a metadata týkající se platnosti modelu, informace o souřadnicovém systému a další informace.



obrázek č.3 – struktura souboru GeoTIFF, který obsahuje více pásem a metadata

## Ověřování zvoleného způsobu ukládání

Pro ověřování zvoleného způsobu ukládání byla využita grafická knihovna GDAL, která umožňuje provádět mnoho operací a rastrovými daty a formáty (čtení, zápis, transformace, konverze a další). Je využívána v mnoha aplikacích (GRASS, QGIS, FWTools, UMN MapServer a další). Obsahuje mnoho programových komponent z nichž jsem pro testování používal zejména gdalinfo pro získávání informací o souborech, gdal\_translate pro provádění konverzí mezi jednotlivými formáty a vkládání metatagů a gdal\_merge pro spojování do souborů do jednoho souboru

Driver: GTiff/GeoTIFF  
Size is 400, 300  
Coordinate System is  
Metadata:

**Základní metadata**

**Base metadata**=<?xml version="1.0" encoding="windows-1250"?><!--Minimal smXML based instance document--> <MD\_Metadata.....</MD\_Metadata>

Chanel 1 Validity=<ValidFrom>2006-04-22 7:00:00</ValidFrom><ValidTo>2006-04-22 7:30:00</ValidTo> <DateStamp>2006-4-24 23:20:31</DateStamp>  
Chanel 2 Validity=<ValidFrom>2006-04-22 7:30:00</ValidFrom><ValidTo>2006-04-22 8:00:00</ValidTo> <DateStamp>2006-4-24 23:22:21</DateStamp>  
Chanel 3 Validity=<ValidFrom>2006-04-22 8:00:00</ValidFrom><ValidTo>2006-04-22 8:30:00</ValidTo> <DateStamp>2006-4-24 23:27:21</DateStamp>  
Chanel 4 Validity=<ValidFrom>2006-04-22 8:30:00</ValidFrom><ValidTo>2006-04-22 9:00:00</ValidTo> <DateStamp>2006-4-24 23:28:11</DateStamp>

**Metadata týkající se platnosti**

Corner Coordinates:

Upper Left ( 0.0, 0.0)  
Lower Left ( 0.0, 300.0)  
Upper Right ( 400.0, 0.0)  
Lower Right ( 400.0, 300.0)  
Center ( 200.0, 150.0)

Band 1 Block=400x20 Type=Byte, ColorInterp=Gray  
Band 2 Block=400x20 Type=Byte, ColorInterp=Undefined  
Band 3 Block=400x20 Type=Byte, ColorInterp=Undefined  
Band 4 Block=400x20 Type=Byte, ColorInterp=Undefined

**Jednotlivé kanály uložené v souboru**

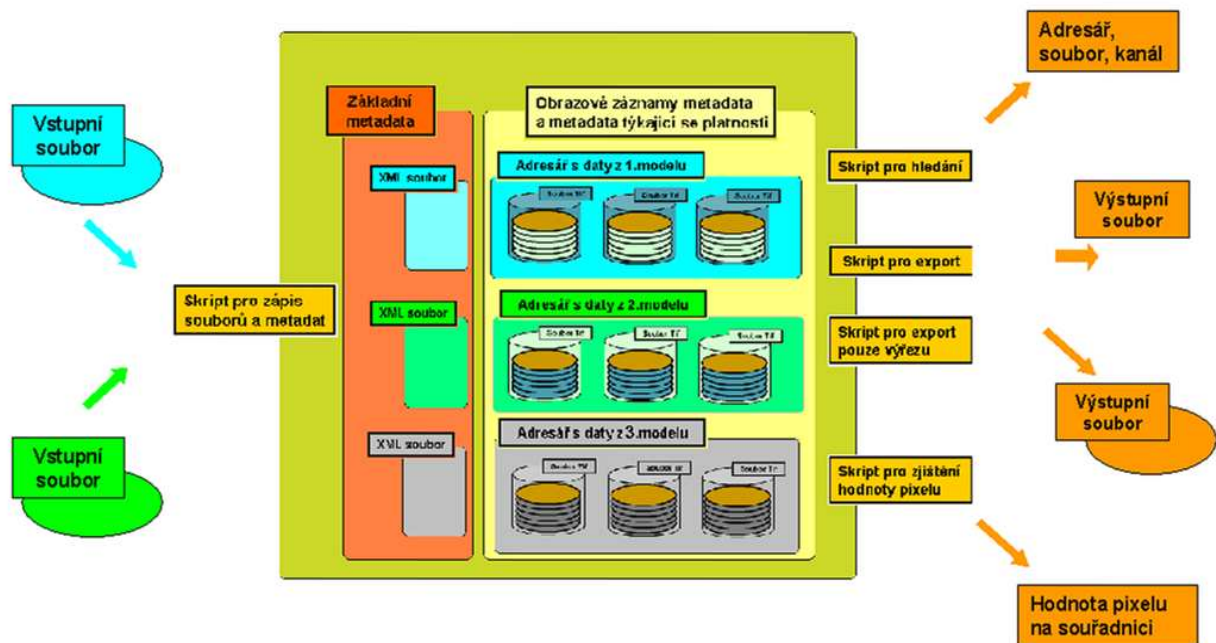
obrázek č.4 – upravený výpis informací o souboru, který obsahuje čtyři pásma

## Návrh automatizace správy dat

Ke knihovně Gdal lze přistupovat přes API a to z programovacích jazyků C/C++, Perl, Python. Jako programovací jazyk byl zvolen jazyk Python, mimo jiné i proto, že jednotlivé programové komponenty jsou také napsány v tomto programovacím jazyce. Bylo navrženo několik objektově orientovaných skriptů, které automatizují celý proces správy rastrových dat.

- Hlavní skript umožňující automatizaci vkládání kanálů a metadat do rastrového souboru
- Skript pro vyhledávání potřebných rastrových souborů a jednotlivých kanálů
- Skript umožňující export potřebného kanálu do samostatného souboru
- Skript umožňující export pouze výřezu z daného obrazového záznamu
- Skript umožňující zjištění hodnoty pixelu na zvolené souřadnici





obrázek č.5 – schéma navrhovaného řešení

## Závěr

Hlavním úkolem této bakalářské práce bylo ověřit možnosti archivace časových sérií rastrových dat, včetně jejich metadat. Jedním z hlavních cílů práce bylo ověření možnosti zápisu metadat primárně do rastrového formátu. Navržené řešení efektivní správy rastrových dat v časových sériích je založeno na využití grafické knihovny GDAL pro práci s rastrovými daty a značkovacím jazykem XML pro ukládání metadat. Výsledkem této práce je také série skriptů zajišťující efektivní správu rastrových dat v časových sériích. Funkčnost jednotlivých skriptů byla otestována na datech z modelu GFS a datech z radarových měření CHMÚ. Všechny úkoly vyplývající ze zadání práce byly tedy splněny v plném rozsahu.

Největší výhodou tohoto řešení spočívá v možnosti použití jakéhokoliv standardu pro metadata ve formátu XML. Dalšími výhodami jsou bezesporu také nezávislost na použité platformě, omezení počtu archivovaných souborů a možnost dalšího rozšíření a možnost implementace skriptů do jiných systémů.

## Seznam informačních zdrojů

- 1 ALADIN. *Aire Limitée, Adaptation Dynamique, Development International* [online]. Dostupné na WWW : <<http://www.chmi.cz/meteo/ov/aladin/>>
- 2 BRABEC, Stanislav. *Grafika v UNIXu - bitmapové formáty TIFF a JPEG*. [online]. Dostupné na WWW: <<http://www.root.cz/clanky/bitmapove-formaty-tiff-jpeg/>>
- 3 DAŘENA, František. *Perl* [online]. Dostupné na WWW: <<http://akela.mendelu.cz/~darena/Perl/perl.html>>
- 4 GAULD, Alan. *Learning to program* [online]. Dostupné na WWW: <<http://www.freenetpages.co.uk/hp/alan.gauld/tutor2/index.htm>>
- 5 GeoTIFF. *GeoTIFF* [online]. Dostupné na WWW: <<http://www.remotesensing.org/geotiff/geotiff.html>>

- 6 GDAL. *Geospatial Data Abstraction Library* [online]. Dostupné na WWW: <<http://www.gdal.org/>>
- 7 HAKL, Aleš. *Programujeme s XML – DOM*. [online]. Dostupné na WWW : <[http://www.linuxsoft.cz/article.php?id\\_article=168](http://www.linuxsoft.cz/article.php?id_article=168)>
- 8 HAKL, Aleš. *Python*. [online]. Dostupné na WWW: <[http://www.linuxsoft.cz/article\\_list.php?id\\_kategorie=217](http://www.linuxsoft.cz/article_list.php?id_kategorie=217)>
- 9 HARMS, Daryl, a MCDONALD, Kenneth. *Začínáme programovat v jazyce Python*. 1. vyd. Brno : Computer Press, 2003. 456 stran. ISBN 80-7226-799-X
- 10 HOLÝ, Jakub. *Grafické formáty*. [online]. Dostupné na WWW: <<http://digi.zive.cz/h/technika/AR.asp?ARI=157>>
- 11 ISO 19 115. Norma ČSN ISO 19 115. Zapůjčená VŠB – TU pro studijní účely
- 12 ISO/PDTS 19139.4. Norma ISO/PDTS 19139.4. Zapůjčená VŠB – TU pro studijní účely
- 13 KOSEK, Jiří. *Seriál o XML pro Softwarové noviny*. [online]. Dostupné na WWW : <<http://www.kosek.cz/clanky/swn-xml/>>
- 14 KOSEK, Jiří. *SGML: Standard Generalized Markup Language*. [online]. Dostupné na WWW : <<http://www.kosek.cz/clanky/cw/sgml.html>>
- 15 KOSEK, Jiří. *XML API*. [online]. Dostupné na WWW : <<http://www.kosek.cz/xml/api/>>
- 16 KRÁČMAR, Jan. *Meteorologické radiolokátory*[online]. Dostupné na WWW: <<http://www.chmi.cz/meteo/rad/index.html>>
- 17 KREJČÍ, Richard. *Encyklopedie publikačních formátů: TIFF*. [online]. Dostupné na WWW: <<http://www.grafika.cz/art/polygrafie/entiff.html>>
- 18 MATOUŠEK, Jan. *Grafické formáty*. [online]. Dostupné na WWW: <<http://grafika.borec.cz/formaty.html>>
- 19 OGD. *Open Geographic Datastore Interface*. [online]. Dostupné na WWW : <<http://ogdi.sourceforge.net/>>
- 20 Orlik A., Ruzicka J., Stromsky J., Dergel P., Kamler J. *Správa časoprostorových dat v prostředí PostgreSQL/Postfix*. . In Sborník z konference GIS Ostrava 2006, Dostupné na WWW: <[http://gis.vsb.cz/gisengl/Conferences/GIS\\_Ova/GIS\\_Ova\\_2006/Proceedings/Referaty/orlik.html](http://gis.vsb.cz/gisengl/Conferences/GIS_Ova/GIS_Ova_2006/Proceedings/Referaty/orlik.html)>
- 21 PAVLOVIČ, Jan. *Tvorba dokumentu v XML*. Bakalářská práce obhájená na Fakultě informatiky Masarykovy univerzity r.2001. [online]. Dostupné na WWW : <<http://www.fi.muni.cz/~xpavlov/xml/examples/bc1/bc1.html>>
- 22 PNG, *Portable Network Graphics* [online]. Dostupné na WWW: <<http://www.libpng.org/pub/png/>>
- 23 PROFANT, Vít. *Vývoj aplikací XML s využitím programovacího jazyka Java*. Bakalářská práce obhájená na Pedagogické fakultě Jihočeské univerzity r.2003.

- [online]. Dostupné na WWW :  
<<http://home.tiscali.cz/ca895221/diplomka/docbook/html/ch01s02.html> >
- 24 PYTHON. *The Python Programming Language* [online]. Dostupné na WWW:  
<<http://www.python.org/>>
- 25 SAROUN, Martin. *Moderní grafické formáty*. [online]. Dostupné na WWW :  
<http://wise.felk.cvut.cz/~apg/apg-tutorials02/ch04s51.html>
- 26 ŠUMBERRA, Stanislav. *Řešení správy rastrových dat s využitím databáze Oracle 9i a MicroStation v8*. In Sborník z konference GIS Ostrava 2003, Dostupné na WWW: <[http://gis.vsb.cz/GIS\\_Ostrava/GIS\\_Ova\\_2003/Sbornik/Referaty/sumbera.htm](http://gis.vsb.cz/GIS_Ostrava/GIS_Ova_2003/Sbornik/Referaty/sumbera.htm)>
- 27 ŠVEC, Jan. Učebnice jazyka Python [online]. Dostupné na WWW:  
<<http://vik.sh.cvut.cz/~tnt/python/tut/>>
- 28 TOMA, Milan. *Základy programovacího jazyka Python* [online]. Dostupné na WWW:<<http://www.toma.ahoj.sk/>>
- 29 VÁVRA, Jan. *LZW komprese*[online]. Dostupné na WWW:  
<<http://xien.jikos.cz/prog/MyLZW.html>>
- 30 VIVODA, Josef. *Numerická predpoveď počasia* [online]. Dostupné na WWW :  
<<http://www.pu.army.sk/rservice.php?akce=tisk&cislocianku=2005032901>>
- 31 WIKIPEDIE, otevřená encyklopedie. *C (programovací jazyk)* Dostupné na WWW:  
<[http://cs.wikipedia.org/wiki/Jazyk\\_C](http://cs.wikipedia.org/wiki/Jazyk_C)>
- 32 WIKIPEDIE, otevřená encyklopedie. *LZW, Lempel-Ziv-Welch* [online]. Dostupné na WWW: <<http://en.wikipedia.org/wiki/LZW>>
- 33 VOJTEK, David. *Úvod do geoinformačních technologií* [online]. Dostupné na WWW : <<http://gis.vsb.cz/vojtek/UGIT/seznam.html>>
- 34 VRÁNEK, Petr. *Školní mapový server*. [online]. Dostupné na WWW :  
<<http://logout.sh.cvut.cz/~vrany/mapsvr.php>>
- 35 W3C. XML Schema Definition [online]. Dostupné na WWW:  
<<http://www.w3.org/TR/xmlschema>>
- 36 XML. *Úvod do XML*. [online]. Dostupné na WWW :  
<<http://www.tvorba-webu.cz/xml.php3>>