

Metodika digitalizace a zpřístupnění starých atlasů

Ondřej Böhm, Filip Antoš, Klára Ambrožová, Jan Havrlant, Milan Talich

Realizováno z programového projektu DF11P01OVV021: Program aplikovaného výzkumu a vývoje národní a kulturní identity financovaného MK ČR v rámci projektu

„Kartografické zdroje jako kulturní dědictví. Výzkum nových metodik a technologií digitalizace, zpřístupnění a využití starých map, plánů, atlasů a glóbulů.“

Výzkumný ústav geodetický, topografický a kartografický, v.v.i.

červen 2015

Obsah

1. Předmět metodiky
2. Struktura metodiky
3. Digitalizace
 - 3.1 Pořízení dat
 - 3.2 Datový model
4. Zpřístupnění na internetu
 - 4.1 Technologie pro zobrazení rastrů na internetu
 - 4.2 Datový model pro zpřístupnění
 - 4.3 Struktura aplikace pro zpřístupnění
 - 4.4 Hardwarové nároky
5. Pro koho je metodika určena
6. Seznam použitých zdrojů
7. Přílohy
 - 7.1 Zápis bounding boxu ve formátu MARC21
 - 7.2 Zápis bounding boxu po částech

1. Předmět metodiky

Staré atlasy jsou součástí našeho kulturního dědictví. Nejen že jsou součástí naší historie a názorně vykreslují situaci v době svého vzniku, ale jsou také v pravém slova smyslu uměleckými díly dokládajícími zručnost a uměleckou vyspělost našich předků.

Přínos digitalizace starých atlasů je zřejmý. Jednak má ochranný účel - místo s originály lze pracovat s jejich digitálními kopiemi. Za druhé umožňuje zpřístupnění atlasů mnohem širšímu okruhu zájemců, ať již odborníkům či laické veřejnosti, navíc často s přidanými funkcemi, jako přibližná či přesná georeference map, snadná navigace v atlasech, provázanost s dalšími dokumenty a aplikacemi a další.

Svým obsahem atlasy stojí na pomezí mapy a knihy a proto jejich digitalizace i zpřístupnění na internetu přináší některé specifické problémy. Tato metodika předkládá způsob jak tyto problémy řešit a efektivně digitalizovat staré atlasy a zpřístupňovat je na internetu. Zvláštní důraz je kladen na datový model pro reprezentaci atlasů a jejich metadat, protože dobře navržený datový model je základem pro archivaci i zpřístupňování digitalizovaných atlasů.

2. Struktura metodiky

Metodika je rozdělena do dvou hlavních částí. První je věnovaná digitalizaci a datovému modelu atlasu pro jeho evidenci a archivaci.

Druhá část je věnována zpřístupňování digitalizovaných atlasů na internetu. Popisuje datový model vhodný pro vystavení atlasů, jeho vztah k modelu pro archivaci a zásady a doporučené technologie pro tvorbu webových aplikací pro prohlížení digitalizovaných atlasů.

3. Digitalizace

Pod pojmem digitalizace rozumíme převod originálu do digitální podoby. Je třeba si uvědomit, že se nejedná pouze o skenování ale také o navazující činnosti, jako opatření naskenovaných souborů metadaty, jejich uložení, případný postprocessing, vytvoření záložních kopií, atd. Důležitou součástí digitalizace je také volba datového modelu, vhodného formátu ukládání dat a metadat a zálohování.

U atlasů je zvláště důležitý datový model, neboť nestačí spojit metadata s daným rastrovým obrazem, ale je nutné také zachytit strukturu a obsah atlasu - pořadí a obsah jednotlivých stránek.

3.1 Pořízení dat

Jelikož je mapový atlas soubor map, je potřeba při jeho digitalizaci dodržovat stejná pravidla jako při digitalizaci map. To znamená brát v potaz jejich vznik a kartografické vlastnosti a zachovat tak při digitalizaci plný informační obsah map. Především je třeba dbát na zachování kartometrických vlastností map v atlasu, tzn. provádět skenování na kartometrickém skeneru, který je zkonstruován tak, aby při digitalizaci nedocházelo ke geometrické deformaci obrazu mapy [1].

Zároveň mají atlasy charakter knihy, což také přináší některá specifika. Je nutné vypořádat se s celkovou tloušťkou a tuhostí předlohy (tzn. atlas se musí do skeneru vejít) a s rozdílnou tloušťkou rozevřeného atlasu.

Výhodným postupem je použít kartometrický skener s kolébkou pro vyrovnání tloušťky vazby a skenovat atlas po dvojicích stránek - vždy celou dvoustranu rozevřeného atlasu jako jeden sken. Kolébka odstraní potíže s výškou vazby a skenování po dvojicích stran vyloučí nutnost spojovat mapy zabírající dvě strany.



Obr. 1 - Kartometrický skener ScannTech 800 a vpravo integrovaná kolébka pro digitalizaci atlasů

Dále je třeba skenovat atlasy s dostatečným rozlišením. Pro vystavení na internetu stačí rozlišení 300 DPI - 400 DPI, ale je lepší skenovat s rozlišením 600 DPI - 800 DPI pro případ navazujícího zpracování map z atlasu v aplikacích jako vyhledávání mapových značek, klasifikace apod.

3.2 Datový model atlasu

Úkolem datového modelu je zachytit strukturu, obsah a vlastnosti atlasu. Je to tedy vhodně uspořádaný a uložený soubor metadat. Tato metadata obsahují informace užitečné pro evidenci atlasu (název, autor, rok vydání atd.), pro rekonstrukci atlasu (číslo stránky, druh obsahu atd.) a doplňující informace (anotace atd.). Součástí návrhu datového modelu je také způsob jeho ukládání.

Základem navrhovaného datového modelu je pohled na atlas jako na soubor menších celků. Hlavní, do jisté míry abstraktní, entitou je *Atlas*. Tato entita reprezentuje atlas jako dílo a obsahuje povšechné informace o atlasu. Jsou to:

- id - unikátní identifikátor atlasu v rámci budovaného systému evidence
- nazev - jméno atlasu
- autor - autor atlasu
- vydal - vydavatel atlasu
- vlastnik - instituce v jejíž sbírce se nachází originál atlasu
- signatura - identifikátor atlasu v rámci evidenčního systému vlastníka
- rok_vydani_presny - přesný rok vydání (je-li znám)
- rok_vydani_min - spodní hranice u orientačního data vydání
- rok_vydani_max - horní hranice u orientačního data vydání
- anotace - popis atlasu

Atlas se dále dělí na *fragmenty*. *Fragment* má již konkrétní význam - každý *fragment* reprezentuje jednu stranu či dvoustranu atlasu nebo jeho obálku. Důvodem pro různou velikost (ve stránkách) jednotlivých *fragmentů* jsou mapy obsažené v atlasu. Ty často zabírají celou dvoustranu a nedává smysl takovou mapu dělit na části. Proto *fragment* zabírá proměnlivý počet skutečných stran atlasu.

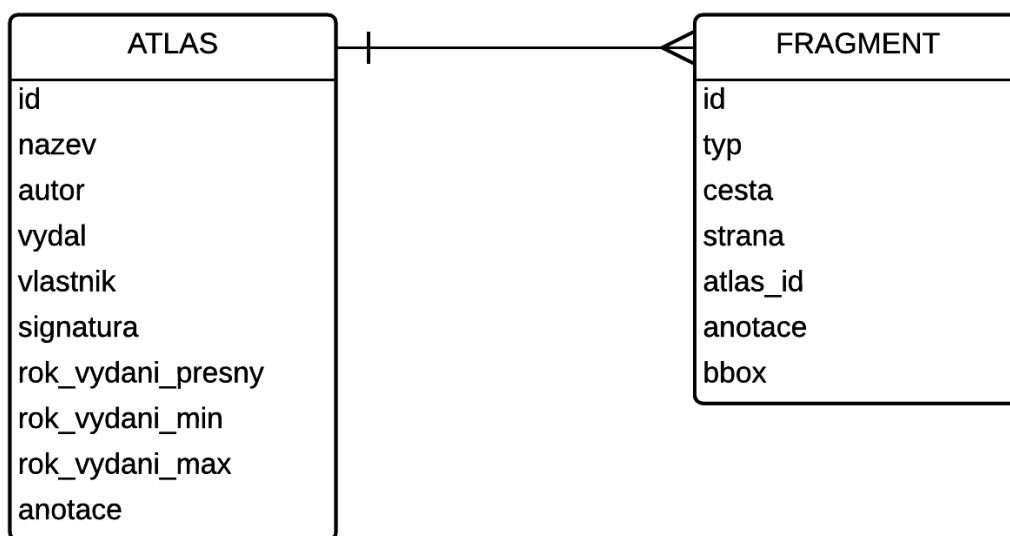
Fragment obsahuje informace vztahující se pouze k dané části atlasu:

- id - unikátní identifikátor fragmentu
- typ - druh obsahu. Může nabývat hodnot: *obalka*, *text*, *mapa*, *mapa+text*
- cesta - cesta k rastrovému obrazu fragmentu (skenu)
- strana¹ - pořadí fragmentu v atlasu. 0 pro obálku
- atlas_id - id atlasu do kterého fragment patří
- anotace - popis fragmentu
- bbox - přibližné geografické souřadnice, *Bounding Box* [8], ohraničující zobrazené území, zapsané ve formátu MARC21 (viz příloha 1). Prázdné pro fragmenty typu *obalka*, *text*.

Za povšimnutí stojí pole *bbox*. To je prázdné pro fragmenty typu *obalka* a *text*, ale u ostatních hraje velmi důležitou roli, protože více či méně přesně určuje, jaké území mapa zobrazuje.

Způsobů jak metadata ukládat je celá řada. Velmi dobře se k tomu hodí relační databáze. Ta přirozeně reflektuje vztah mezi atlasem a jeho fragmenty. Schéma takové databáze je na obr. 2.

¹ Toto číslo vyjadřuje pořadí stránky v atlasu, nemusí souhlasit s číslem vytištěným na stránce (jsou-li stránky číslovány)



Obr. 2 - schéma databáze pro archivaci atlasů

4. Zpřístupnění modelů atlasů na internetu

Zpřístupnění digitalizovaných atlasů na internetu vyžaduje vyřešit následující problémy:

1. Zvolit vhodný formát pro zpřístupnění rastrových dat
2. Navrhnout vhodný datový model pro reprezentaci atlasu
3. Vytvořit aplikaci pro prohlížení atlasů
4. Zajistit server s dostatečným výkonem a konektivitou pro obsluhu příchozích požadavků

4.1 Formát rastrových dat pro zpřístupnění

Skeny stránek atlasů jsou obecně příliš velké, než aby bylo možno je vystavit na internetu vcelku - to by bylo příliš náročné na datové přenosy a na výkon internetového prohlížeče. V neposlední řadě by takový způsob poskytoval pramalý uživatelský komfort. Naštěstí dnes existuje řada způsobů jak na internetu zobrazovat velká rastrová data.

Ideální je použít formát Zoomify [2] - původní rastr se rozřeže na dlaždice 256 x 256 pixelů v několika stupních zoomu. Výsledné dlaždice jsou uspořádány do adresářové struktury podle jejich polohy v celkovém obraze [3]. Toto řešení má několik zásadních výhod:

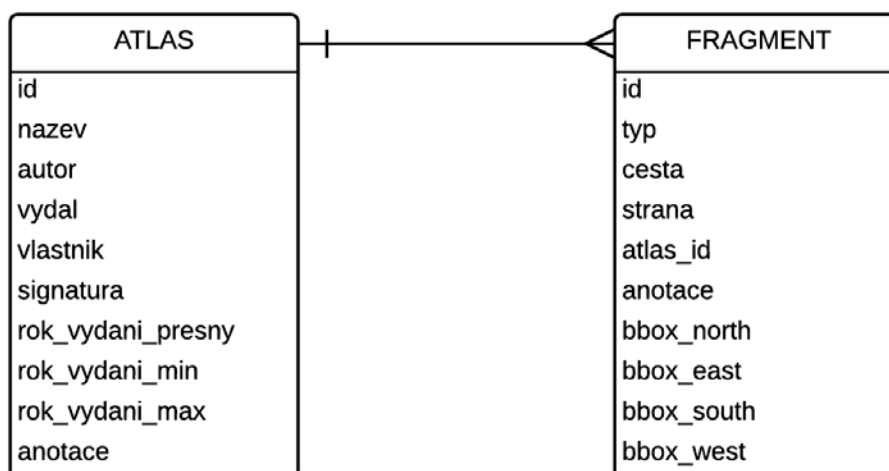
- zmenšení velikosti dat (kompresce při ukládání dlaždic významně sníží celkovou velikost obrazu při zachování dostatečné kvality pro prohlížení na monitoru),
- jednoduché poskytování výsledku - dlaždice jsou obyčejné obrázky, které lze poskytovat libovolným webovým serverem s minimálními nároky na výkon. Není potřeba žádného speciálního software ani složitého, či výpočetně náročného zpracování požadavků na jednotlivé dlaždice,
- program na konverzi rastrů do formátu Zoomify je dostupný zdarma.

4.2 Datový model pro zpřístupnění atlasů

Samotná obrazová data jsou ale pouze dílčí částí zpřístupnění. Stejně jako v případě digitalizace hraje zásadní úlohu datový model atlasu a způsob uspořádání a uložení metadat včetně jejich provázání s obrazovými daty. Zde vyjdeme z modelu atlasu představeného v předchozí kapitole a minimálními úpravami dostaneme model vhodný pro použití v aplikacích zpřístupňujících digitalizované atlasy (viz obr. 3). Tento upravený model je navržen s dvěma hlavními cíli:

1. věrná reprezentace struktury a obsahu atlasu,
2. snadná navigace v rámci atlasu (zobrazení konkrétní stránky atd.),
3. možnost snadného hledání map zobrazujících území o známých geografických souřadnicích nejen v rámci atlasu, ale také napříč všemi atlasy.

První a druhá podmínka je splněna již v původním "archivním" modelu - relace mezi atlasem a fragmenty společně se sloupci *stranka_min* a *stranka_max* postihují strukturu atlasu.



Obr. 3 - schéma databáze pro zpřístupnění atlasů

Pro splnění třetí podmínky rozdělíme pole *bbox* na čtyři samostatná pole *bbox_north*, *bbox_east*, *bbox_south* a *bbox_west* obsahující zeměpisné souřadnice ve stupních (viz příloha 2). To usnadní a urychlí vyhledávání ve fragmentech atlasu podle polohy.

A konečně pole *cesta* v novém modelu nebude obsahovat cestu k obrazovému souboru (skenu), ale k adresáři obsahujícímu zoomify reprezentaci tohoto rastru.

Pro implementaci nového modelu je opět ideální relační databáze jako přirozený způsob vyjádření vztahů mezi atlasem a jeho obsahem.

4.3 Aplikace pro zpřístupnění atlasů

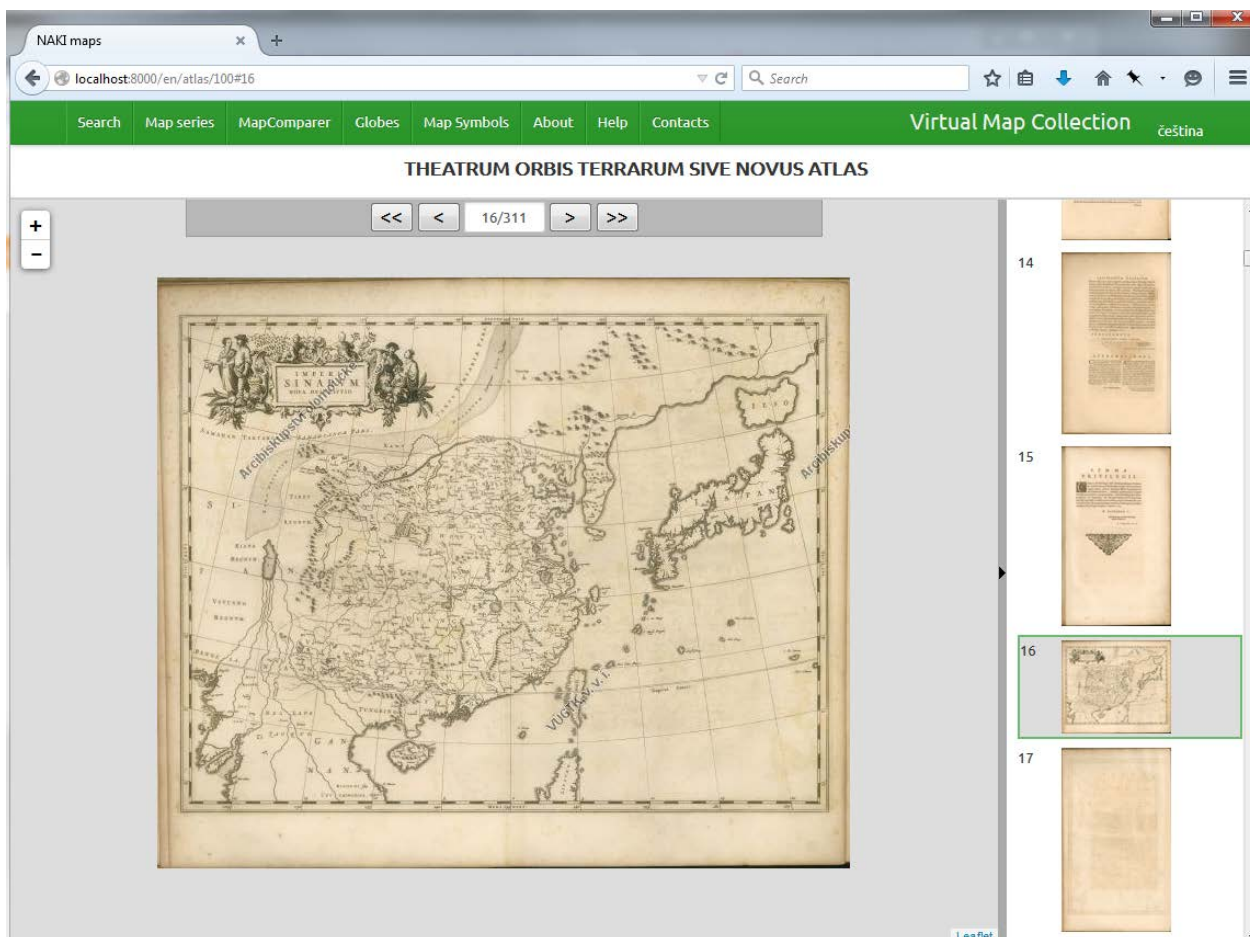
Nyní máme data vhodná pro vystavení na internetu a máme model popisující tato data. Zbývá implementace aplikace, která tyto dvě komponenty zužitkuje a umožní prohlížení atlasů online.

Konkrétní návrh a podoba takové aplikace by měla vycházet ze specifik konkrétní sbírky atlasů, účelu ke kterému má sloužit (např. zda jde o samostatnou aplikaci, nebo o komponentu většího systému) a projektu pro který vzniká. Obecně ale můžeme říci, že taková aplikace by měla obsahovat dvě základní funkcionality:

1. prohlížečku digitalizovaných atlasů,
2. vyhledávací nástroje.

První bod je jasný a vyplývá ze samé podstaty zadání. Implementace takové prohlížečky je dnes poměrně jednoduchá díky knihovnám jako Leaflet [4] s pluginem ZoomifyLayer [5] nebo OpenLayers 3 [6]. Ty umožňují vytvořit přívětivé uživatelské rozhraní pro zobrazení rastru ve formátu Zoomify včetně funkcí pro posouvání, přibližování a oddalování.

Prohlížečka by měla také obsahovat nástroje pro navigaci v atlasu: seznam s odkazy na jednotlivé stránky nebo možnost přejít přímo na konkrétní stránku, popř. kombinaci obojího.



Obr. 4 - Příklad prohlížečky atlasů (atlas THEATRUM ORBIS TERRARUM SIVE NOVUS ATLAS M. Martina z r. 1655). Uprostřed zobrazovací pole s navigačním panelem v horní části, vpravo seznam stránek.

Smysl vyhledávacího nástroje je také zřejmý. V případě skutečně malých sbírek jeho funkci snadno zastane obyčejný seznam atlasů. Ale v případě větších sbírek je nutné mít možnost jednoduše a rychle vyhledat konkrétní atlas který uživatele zajímá, nebo atlasy splňující určitá kritéria (autor, rok vzniku, atd.).

4.4 Hardwarové zabezpečení

Výhodou předkládaného řešení je jeho univerzálnost. Nevyžaduje žádný speciální software dostupný pouze pro některé platformy, i když do jisté míry záleží na technologii zvolené k vývoji prezentační aplikace.

Je ale třeba mít na paměti relativní datovou náročnost - digitalizované atlasy jsou obrazová data a i ve formátu zoomify mohou představovat značnou zátěž na server a přenosovou kapacitu jeho připojení, zejména při větším počtu přístupů. Je také třeba počítat s dostatečnou velikostí úložiště pro digitalizované atlasy úměrnou počtu atlasů.

5. Pro koho je metodika určena

Uplatnění metodiky je především v paměťových institucích, jako jsou archivy, muzea nebo knihovny, které vlastní staré atlasy a mají zájem je zpřístupnit široké veřejnosti ke studiu online prostřednictvím internetu. Dále pak organizacím, které se zabývají digitalizací historických sbírek kartografických děl.

6. Seznam použitých zdrojů

- [1] TALICH M.: Trendy výzkumu možností využívání starých map digitálními metodami. Kapitola v knize: Krajina jako historické jeviště. K počtě Evy Semotanové. Praha : Historický ústav, 2012 - (Chodějovská, E.; Šimůnek, R.), s. 373-386, ISBN 978-80-7286-199-6
- [2] Zoomify - zoomable web images [online] <http://www.zoomify.com/>
- [3] Přidal P., *Zoomify Tile Structure* [online] <<http://www.staremapy.cz/zoomify-analyza/>>
- [4] Leaflet - a JavaScript library for mobile-friendly maps [online]. Vladimir Agafonkin, <<http://leafletjs.com/>>
- [5] ZoomifyLayer - display Zoomify images in Leaflet, [online], O. Böhm < <https://github.com/kalse/zoomify-layer>>
- [6] OpenLayers 3 - *A high-performance, feature-packed library for all your mapping needs*, [online], Open Source Geospatial Foundation < <http://openlayers.org/>>
- [7] MARC21 - Coded Cartographic Mathematical Data [online] <<http://www.loc.gov/marc/authority/concise/ad034.html>>
- [8] Bounding Box [online] <http://wiki.openstreetmap.org/wiki/Bounding_Box>
- [9] Antoš F., Talich M., Böhm O., Havrlant J., Ambrožová K., Soukup L.: *Virtuální mapová sbírka Chartae-Antiquae.cz – důležitý výsledek projektu Kartografické zdroje jako kulturní dědictví*, INFORUM 2014: 20. ročník konference o profesionálních informačních zdrojích, Praha 27.-28. května 2014 [online]. Praha: Albertina icome Praha, 2014.. ISSN 1801–2213, < <http://www.inforum.cz/pdf/2014/antos-filip.pdf>>

7. Přílohy

7.1 Zápis bounding boxu ve formátu MARC21

MARC21 je datový formát vyvinutý původně v Knihovně Kongresu USA pro záznam bibliografických údajů ve strojově čitelné formě. Formát definuje množství polí pro záznam různých informací. Pro bounding box je určeno pole 034 (viz [7]). Příklad zápisu bounding boxu vypadá:

```
$$dE0142428$$eE0145326$$fN0491644$$gN0485050
```

kde znak \$ odděluje jednotlivé položky pole a kód ve tvaru \$<písmeni> určuje význam položky:

- \$d - Souřadnice - západní hranice (maximální zeměpisná délka) území
- \$e - Souřadnice - východní hranice (minimální zeměpisná délka) území
- \$f - Souřadnice - severní hranice (maximální zeměpisná šířka) území
- \$g - Souřadnice - jižní hranice (minimální zeměpisná šířka) území

Souřadnice jsou zapsány ve formátu: *psssmmvv* (polokoule - stupně - minuty - vteřiny), tzn. E0142428 je 14° 24' 28" východní délky.

7.2 Zápis bounding boxu po částech

Pro zpřístupnění atlasů je výhodnější zapsat bounding box po jednotlivých částech (sever, východ, jih, západ) a pouze ve stupních. Bounding box zapsaný ve formátu MARC21 jako

```
$$dE0142428$$eE0145326$$fN0491644$$gN0485050
```

tak pro zpřístupnění zapíšeme do čtyř polí *bbox_north*, *bbox_east*, *bbox_south* a *bbox_west* jako

```
bbox_north: 49,2789  
bbox_east - 14,8906  
bbox_south- 48,8472  
bbox_west - 14,4078
```