

Ověřená technologie 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
glóbů.“

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

červen 2015

Obsah

1. Předmět ověřené technologie
2. Struktura ověřené technologie
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. Ověření technologie
 - 5.1 Vstupní data
 - 5.2 Digitalizace
 - 5.3 Vystavení atlasů
6. Závěr
7. Seznam použitých zdrojů

1. Předmět ověřené technologie

Předkládaná technologie popisuje způsob digitalizace starých atlasů a vystavení digitalizovaných starých atlasů na internetu ve vysoké kvalitě. Dále popisuje způsob uložení metadat o atlasech, principy webové aplikace pro jejich prohlížení a demonstruje tyto na konkrétních příkladech.

Cílem technologie je zpřístupnit staré atlasy široké odborné i laické veřejnosti a přispět tak k popularizaci těchto vzácných památek a k jejich ochraně.

Digitalizované atlasy přináší i některé výhody pro badatele, například snadné prohlížení několika listů současně, v případě georeferencovaných mapových částí snadné porovnání s dalšími mapovými podklady ať již současnými nebo starými.

Tato technologie vychází z Metodiky digitalizace a zpřístupnění starých atlasů [10].

2. Struktura ověřené technologie

Popis technologie je rozdělen do tří hlavních částí. První část se věnuje problematice digitalizace atlasů a popisuje obecné zásady a doporučení ohledně přístrojového vybavení a ukládání dat a metadat.

Druhá část se zabývá zpřístupněním atlasů na internetu v obecné rovině. Věnuje se možnostem vystavení a obecným principům prohlížečky digitalizovaných atlasů.

Třetí část popisuje ověření technologie na konkrétním příkladu tří atlasů. Popisuje způsob skenování atlasů ve VÚGTK, v.v.i. a webovou aplikaci pro prohlížení digitalizovaných starých atlasů, která je součástí mapového portálu Chartae Antiquae (<http://chartae-antiquae.cz>).

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 atlase, 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

Je také 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

Datový model slouží k popisu struktury, obsahu a vlastností atlasu. Je to vhodně uspořádaný a uložený soubor metadat, která 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 atlase. Ty často zabírají celou dvoustranu a nedává smysl takovou mapu dělit na části. Proto *fragment* zabírá proměnlivý počet skutečných stran atlasu.

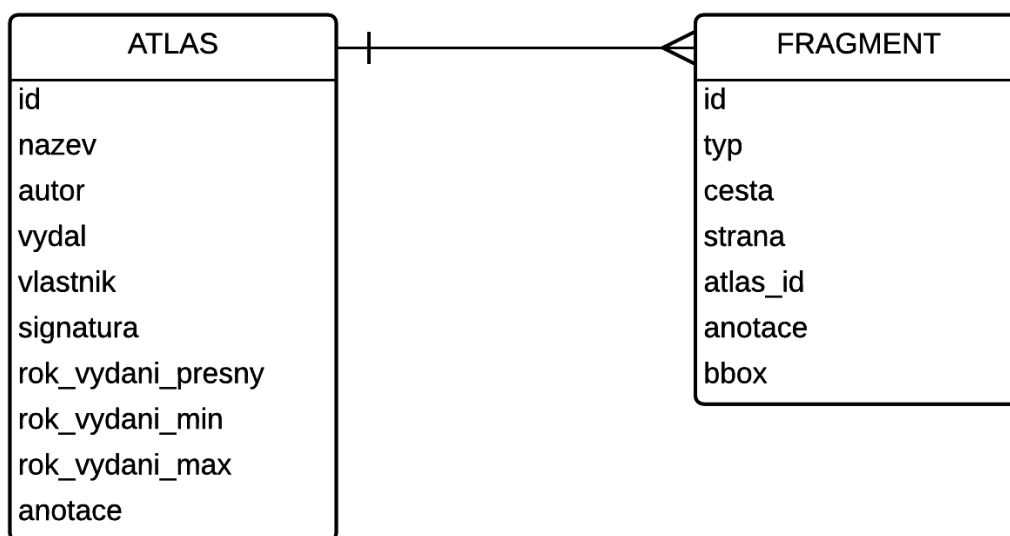
Fragment obsahuje informace vztahující se pouze k dané části atlasu:

- 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.

Metadata lze ukládat různými způsoby. Jedním z nejlepších je relační databáze, protože přirozeně reflektuje vztah mezi atlasem a jeho fragmenty. Schéma takové databáze je na obr. 2.

¹ Toto číslo vyjadřuje pořadí stránky v atlase, 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.

Velmi jednoduchým a přitom velmi dobrým řešením je formát Zoomify [2] - původní rastr se rozřeže na dlaždice 256 x 256 pixelů v několika stupních zoomu. Tyto dlaždice jsou uspořádány do adresářové struktury podle jejich polohy v celkovém obraze [3]. Toto řešení má několik zásadních výhod:

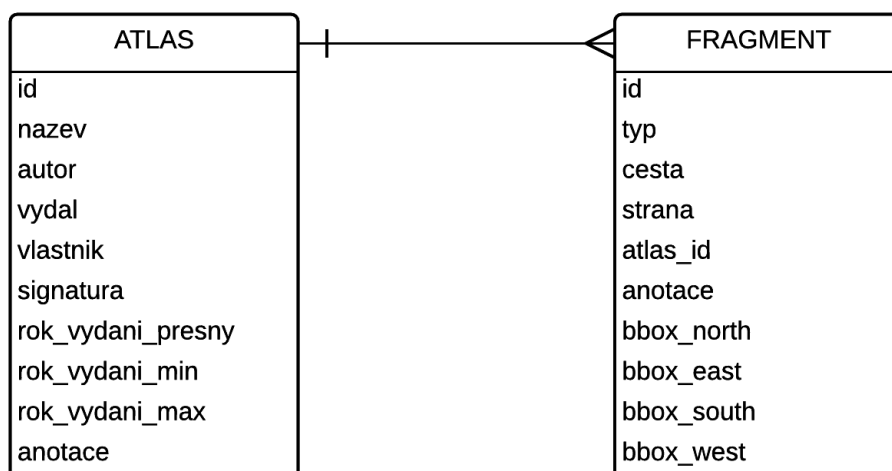
- zmenšení velikosti dat (komprese při ukládání dlaždic významně sníží celkovou velikost obrazu při zachování dostatečné kvality pro prohlížení na monitoru),
- jednoduché poskytování výsledku - dlaždice jsou obyčejné obrázky, které lze poskytovat libovolným webovým serverem s minimálními nároky na výkon. 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. Podobně jako v případě digitalizace hraje i zde 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 se zřetelem na tři hlavní cíle:

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 - vztah mezi atlasem a fragmenty a sloupec *strana* dostatečně postihují strukturu atlasu.



Obr. 3 - schéma databáze pro zpřístupnění atlasů

Pro splnění třetí podmínky je pole *bbox* rozděleno 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.

Pole *cesta* v novém modelu již neobsahuje adresu původního rastru (skenu) v souborovém systému, ale URL adresáře obsahujícího 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.

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 jejich počtu.

5. Ověření technologie

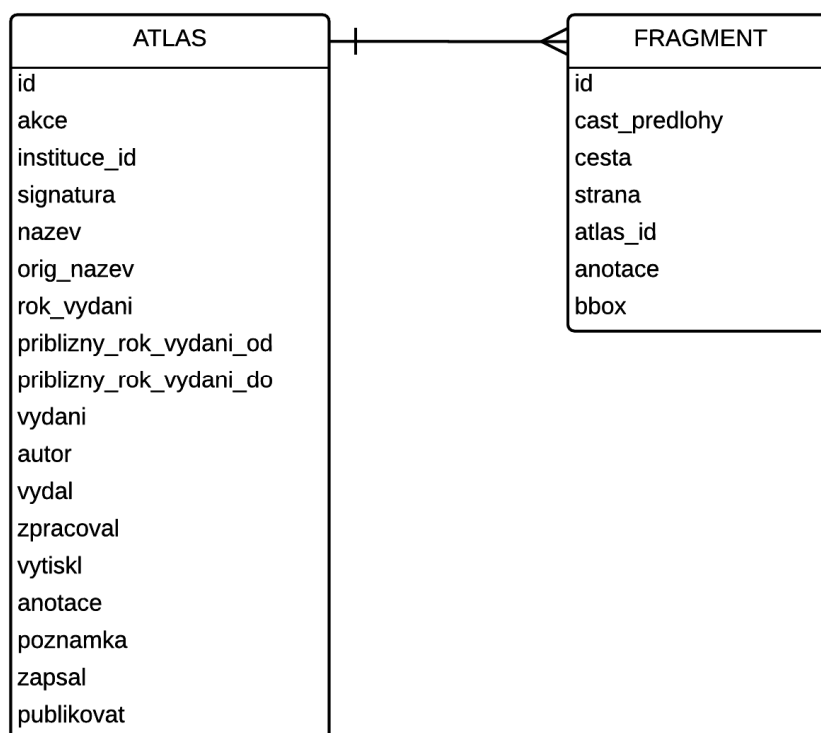
5.1 Vstupní data

Technologie byla ověřena na třech atlasech: THEATRUM ORBIS TERRARUM SIVE NOVUS ATLAS, Martina Martiniho ze sbírky Arcibiskupství olomouckého, atlasu Fredericka de Wita a atlasu Nicolase Sansona ze sbírky Královská kanonie premonstrátů na Strahově.

Všechny atlasy jsou přístupné na internetu na adrese <http://chartae-antiquae.cz/cs/atlas/>, kde jsou také další podrobnosti o nich.

5.2 Digitalizace

Atlasy byly skenovány v rozlišení 600 DPI (atlasy z Královské kanonie premonstrátů na Strahově) a 800 DPI (atlas z Arcibiskupství olomouckého) po jednotlivých dvoustranách na skeneru ScannTech 800 s použitím speciální kolébky pro skenování předloh s charakterem knihy (viz obr. 1). Atlasy i jednotlivé skeny byly opatřeny metadaty uloženými v databázi MySQL pro archivační účely. Schéma databáze použité pro uložení atlasů je na obr. 4.

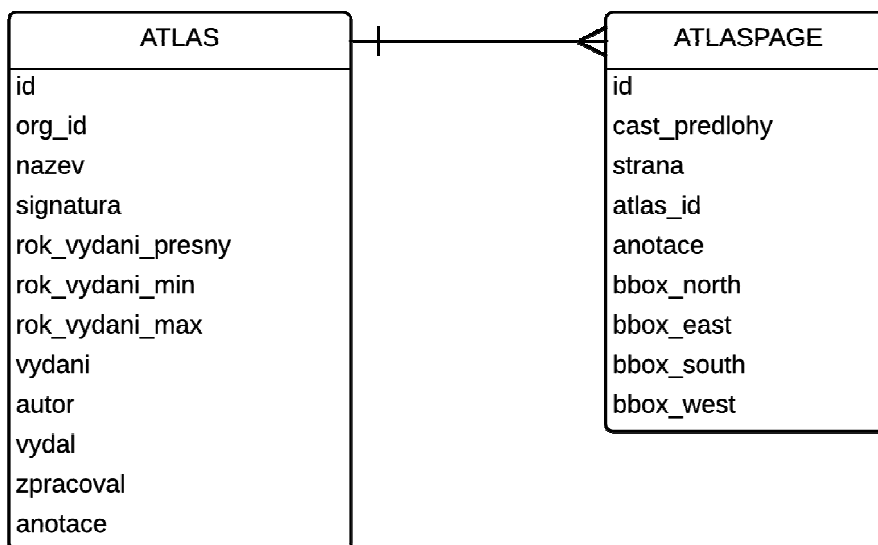


Obr. 4 - Schéma tabulek pro uložení atlasů v archivační databázi

5.3 Vystavení atlasů

Pro vystavení atlasů byly dvoustrany rozřezány na jednotlivé stránky, vyjma případů, kdy celou dvoustranu zabírala jedna mapa. Tyto stránky byly poté převedeny do formátu Zoomify pomocí oficiální Zoomify aplikace.

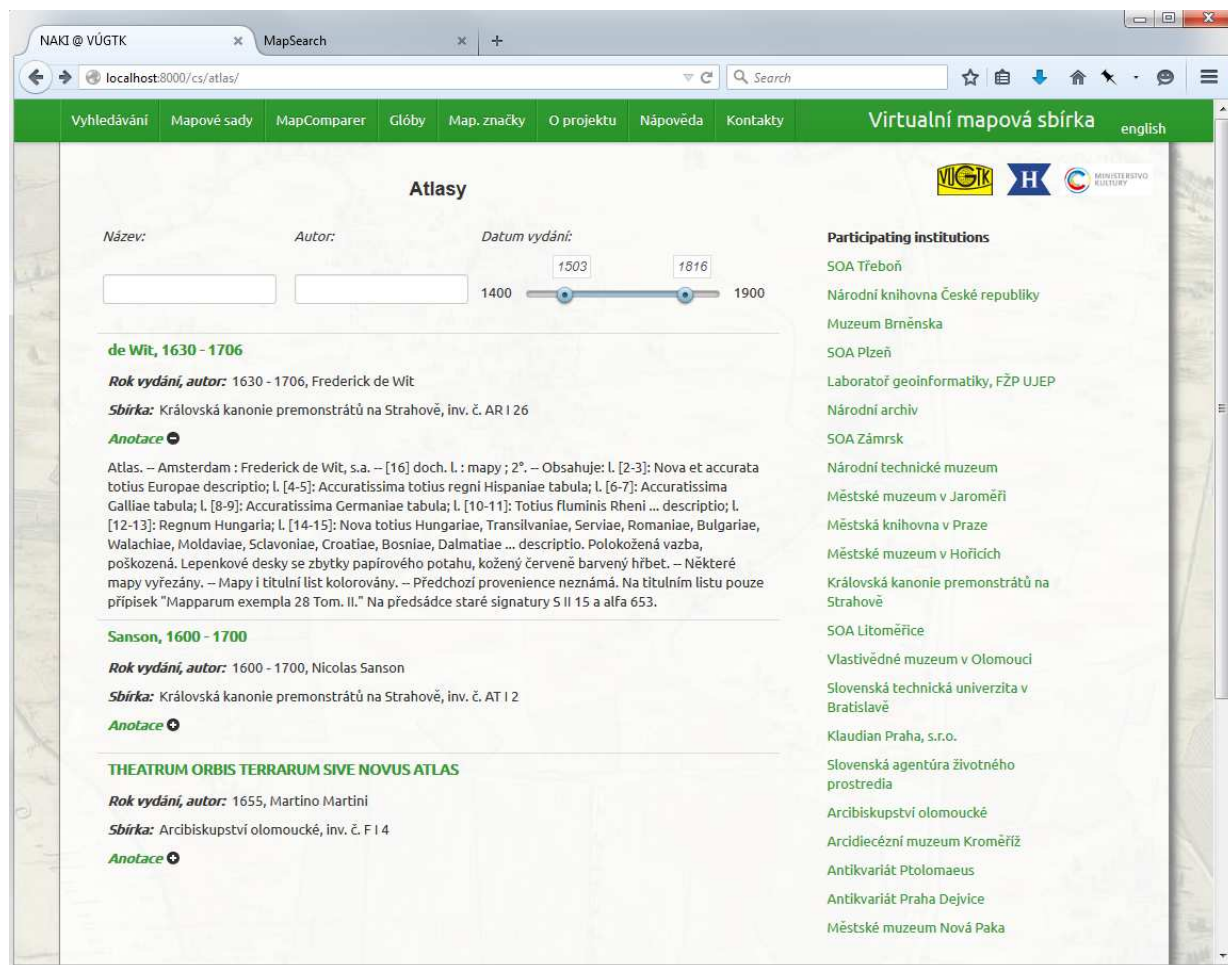
Metadata o atlasech a jednotlivých stránkách byly v souladu s Metodikou digitalizace a zpřístupnění starých atlasů [10] překonvertovány a uloženy do MySQL databáze na serveru Chartae Antiquae. Struktura databáze pro uložení informací o atlasech je na obr. 5.



Obr. 5 - Schéma tabulek pro uložení atlasů v databázi na portálu Chartae Antiquae

Pro prohlížení atlasů byla vytvořena webová aplikace a zakomponována do mapového portálu Chartae Antiquae. Stejně jako ostatní aplikace na portálu je i prohlížečka atlasů postavena na technologiích: webový framework *Django*, javascriptové knihovny *Leaflet* a *Backbone.js* a databáze *MySQL*. Rastry jsou vystavené ve formátu *zoomify*.

Aplikace pro prohlížení atlasů má dvě části. První z nich je seznam vystavených atlasů (viz obr. 6).



Obr. 6 - Seznam atlasů v prohlížečské aplikaci. První z atlasů (de Wit, 1630 - 1706) je se zobrazenou anotací, u ostatních je anotace skryta pro lepší přehlednost

V tomto seznamu jsou uvedeny všechny vystavené atlasy, jejich autoři, roky vydání a anotace. V horní části je formulář pro vyhledávání v atlasech (viz obr. 7) podle názvu, autora a roku vydání. Jméno atlasu slouží jako odkaz na atlas v prohlížeči (viz obr. 8), která tvoří druhou část aplikace.

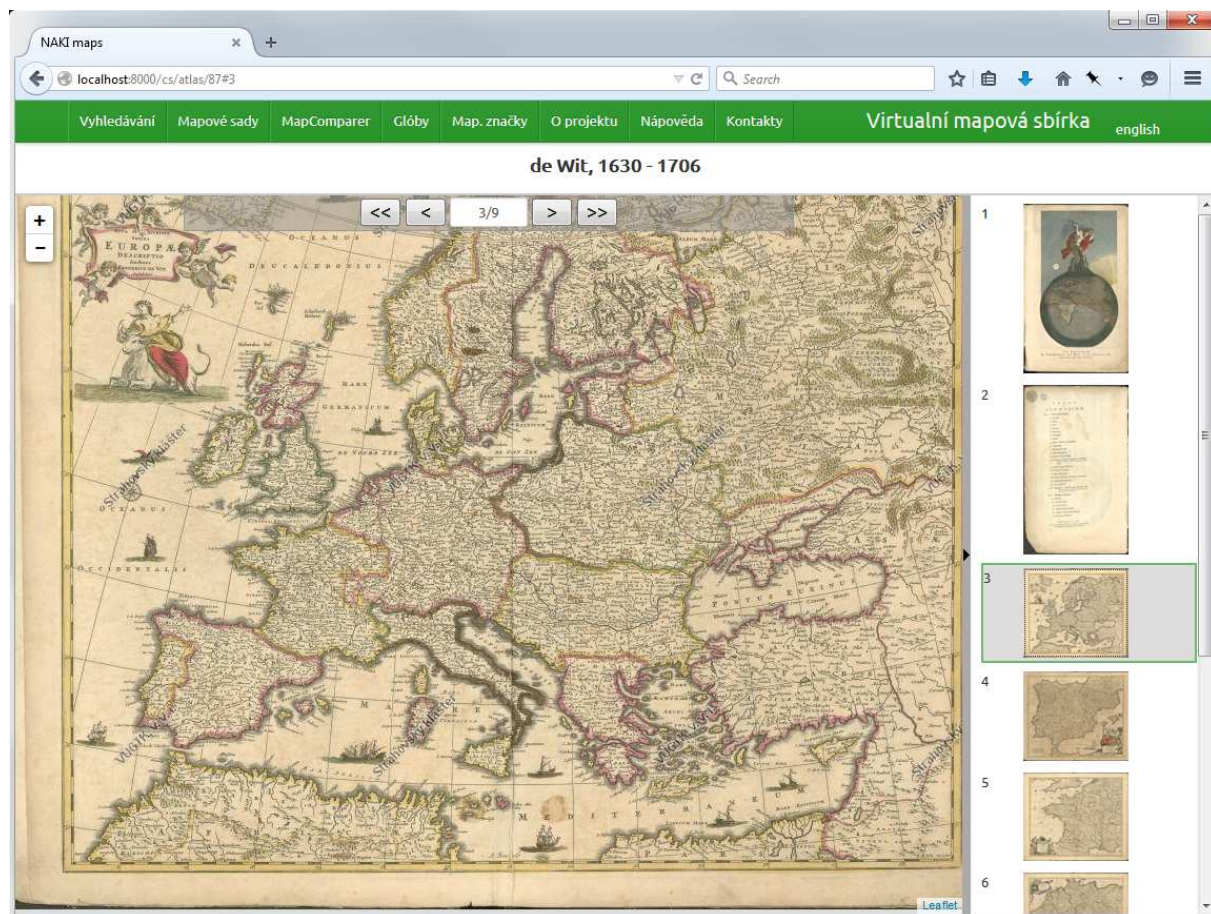


Obr. 7 - Detail vyhledávacích nástrojů

Prohlížečka atlasů (viz obr. 8) se skládá ze tří částí:

- prohlížečské okno
- seznam stránek
- navigační panel

Prohlížečské okno je hlavní částí prohlížečky. Zobrazuje jednu stránku ve formátu zoomify pomocí javascriptové knihovny Leaflet. Dokument je možné posunovat a měnit úroveň zoomu až na plné rozlišení, ve kterém byl atlas skenován.



Obr. 8 - Prohlížečka atlasů

Seznam stránek v pravém postranním panelu ukazuje všechny stránky v atlasu. Aktuální stránka je zvýrazněna ohraničením a změnou barvou pozadí. Jednotlivé stránky v seznamu fungují jako odkazy - po klepnutí na některou z nich je tato stránka zobrazena v prohlížečském okně.

Navigační panel (viz obr. 9) je umístěn v horní části prohlížečského okna. V textovém poli ukazuje číslo aktuální stránky a celkový počet stran atlasu. Číslo stránky je možné přepsat a potvrzením (stiskem klávesy *Enter*) zobrazit tuto novou stránku.



Obr. 9 - Detail navigačního panelu

Kromě toho navigační panel obsahuje čtyři navigační tlačítka s funkcemi (v pořadí zleva doprava) zobrazení první stránky, posun o jednu stránku zpět, posun o jednu stránku vpřed, zobrazení poslední stránky.

6. Závěr

Výše popsaným postupem digitalizace a zpřístupnění tří atlasů byla ověřena funkčnost vyvinuté technologie. Jedná se o atlas THEATRUM ORBIS TERRARUM SIVE NOVUS ATLAS ze sbírky Arcibiskupství olomouckého (Zámek Kroměříž) a dva atlasy Fredericka de Wita a Nicolase Sansona ze sbírky Královské kanonie premonstrátů na Strahově, se kterou byla uzavřena ve smyslu ustanovení § 11 odst. 1 písm. a) zákona č. 130/2002 Sb., o podpoře výzkumu, experimentálního vývoje a inovací z veřejných prostředků a o změně některých souvisejících zákonů (zákon o podpoře výzkumu a vývoje), smlouva o využití výsledků výzkumu a vývoje.

7. 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>>
- [10] Böhm O., Antoš F., Ambrožová K., Havrlant J., Talich M.: *Metodika digitalizace a zpřístupnění starých atlasů*, Zdíby, 2015