



Průběžná zpráva za rok 2010



Brno, 15.11.2010

A Konstatační část

A.1 Úvod

Tato zpráva je třetí roční zprávou projektu Staré mapy online a navazuje na tak zprávy z předešlých let řešení projektu. V úvodu proto jen zopakujeme hlavní cíle projektu:

"Vývoj technologie umožňující formou poloprovozu paměťovým institucím zpřístupňovat na internetu kvalitně georeferencované mapy a další dokumenty grafické povahy. Bude mít tyto vlastnosti:

- umožní uživatelům co nejpřesnější online georeferenci předloh velkých rozměrů
- umožní vyhledávání digitalizovaných map a dalších grafických dokumentů s vazbou k určitému místu s komfortním využitím geografických a časových informací (označení území na mapě, časová osa, map ranking)
- bude dostupná i v anglické verzi a lokalizovatelná do jiných jazyků

Vypracování metodiky pro online zpřístupňování starých map (velkých rozměrů) a dalších grafických dokumentů a jejich georeferencování paměťovými institucemi. Jednotlivé softwarové nástroje vyvinuté v rámci řešení projektu a použitelné i samostatně budou uvolněny pod některou open source licenci."

Řešení tohoto projektu probíhá nadále v kontextu dalších velkých projektů, které MZK řeší nebo se na jejich řešení podílí. Jednak jde o běžící výzkumný záměr "Historické fondy Moravské zemské knihovny v Brně", končící v roce 2010 a pak v květnu 2009 zahájený dvouletý projekt eContentPlus EuropeanaTravel, v jehož rámci je digitalizována Mollova sbírka map.

A.2 Stávající stav

Díky projektu EuropeanaTravel je k dispozici dostatečné množství digitalizovaných starých map. Veškeré práce v rámci tohoto projektu se proto soustředily na analýzy a zejména vývoj potřebného softwaru, mezinárodní spolupráci a prezentaci výsledků vývoje na odborných akcích u nás i ve světě.

A.3 Spolupráce, další projekty

Díky intenzivní propagaci dosavadních výsledků na relevantních odborných fórech se daří výsledky projektu dostávat do povědomí mezinárodní odborné veřejnosti. Zároveň se do sféry našeho zájmu dostávají další projekty.

A.3.1 IPIImage

Na základě specifikací vniklých v rámci našeho projektu byla vyvinuta verze IPIImage, podporující formát Jpeg2000. Vývoj této verze není ještě ukončen, již nyní je však zřejmé, že vývoj tohoto image serveru přišel v pravý okamžik v kontextu rozhodnutí NK a MZK používat Jpeg2000 jak pro archivaci, tak pro zpřístupnění digitálních dokumentů.

V průběhu roku se objevilo již několik zájemců o tuto aplikaci, své využití tak našla mimo jiné při zpřístupňování skenů biologických vzorků na Masarykově univerzitě nebo naposledy v Rakouské národní knihovně.

V rámci končícího výzkumného záměru MZK byla ještě do image serveru doplněna podpora pro standard DeepZoom, volitelná podpora pro viditelný watermarking jednotlivých dlaždic a také podpora pro Memcached. Díky Memcached je možné sdílet již vytvořené dlaždice mezi více běžícími vlákny image serveru a v případě potřeby zároveň i dalšími aplikacemi. Memcached je také možné použít pro sdílení (v našem případě dlaždicové) cache mezi více počítači, v momentě kdy je nutné nasadit image server na více serverů obsluhující velmi často navštěvovanou stránku. Touto formou je Memcached nasazen na serverech projektů jako Wikipedia, Flickr, Twitter, YouTube apod. Díky zapracování podpory MemCached došlo k dalšímu zvýšení rychlosti image serveru a také se otevřeli možnosti jak lépe škálovat celý systém sdílení obrazových dat.

Další projekt MZK navíc řeší i integraci IIPImage do systému Kramerius verze 4, která je v rámci tohoto projektu dále rozšiřována tak, aby bylo možné prostřednictvím Krameria zpřístupňovat mapy a zároveň aby v Krameriu zpřístupněné mapy byly integrovatelné do sady nástrojů vyvinuté v rámci projektu Staré mapy online.

A.3.2 Projekt EuropeanaTravel

MZK se účastní dvouletého digitalizačního projektu EuropeanaTravel v rámci programu eContentplus, který skončí v dubnu 2011. Díky tomuto projektu byla již zdigitalizována Mollova mapová sbírka i další staré mapy a vybrané atlasy z fondů knihovny. V případě sběratelských atlasů jsou bibliograficky popisované jednotlivé mapy coby přívazky, čímž se usnadňuje zpřístupnění map.

V současné době probíhají práce na jejich zpřístupnění, v jehož rámci budou také plně nasazeny nástroje vyvinuté v projektu OldMapsOnline. V rámci projektu Staré Mapy Online byl proto připraven grafický návrh stránek, které budou tuto mapovou sbírku prezentovat. Grafický návrh je v době psaní tohoto textu přístupný na adrese <http://www.kacerr.cz/mzk/> a implementován bude na adrese <http://mapy.mzk.cz>.

A.3.3 Projekt EuropeanaConnect

Projekt EuropeanaConnect byl zahájen v květnu 2009 a bude končit v říjnu 2011. Projekt mimo jiné řeší i implementaci geo-nástrojů do portálu Europeana. Většina prací v této oblasti je soustředěna v AIT (Austrian Institute of Technology), který je tak pro nás velmi perspektivním partnerem. Podrobnosti o tomto projektu jsou uvedeny ve zprávě ze služební cesty na workshop "Place, Cultural Heritage and the Internet" v příloze této zprávy.

A.4 Standardy

Díky rostoucí podpoře standardu W3C HTML5 ve webových prohlížečích je možné začít uvažovat i o praktickém využití jeho prvku canvas, kterým jsme se v našem projektu zabývali už loni a vytvořili jsme první prototyp renderování map s transformací přímo v prohlížeči: <http://examples.maptiler.org/map-reprojection-html5-canvas/>. Na naše úsilí v této oblasti pak navázal další výzkum vývojářů projektu OpenLayers zastřešený Švýcarskou akademickou obcí. Korektně nás citují v relevantní zprávě:

<http://trac.osgeo.org/openlayers/wiki/Future/OpenLayersAndHTML5>

Dalším přirozeným výzkumem v této oblasti je využití kontextu WebGL, který je již podporován ve vývojových verzích mnoha browserů. WebGL je založen na standardu OpenGL ES 2.0 a nabízí programátorské rozhraní pro rychlou 3D grafiku a také pro grafickou kartou akcelerované operace s obrazovými daty.

Důležitým pomocníkem při vývoji multiplatformních webových aplikací je také před rokem zveřejněná knihovna a sada nástrojů Closure (<http://code.google.com/closure/>).

A.5 Jpeg2000

Národní knihovna ČR a Moravská zemská knihovna připravily společný návrh profilu pro jpeg2000, který bude používán v rámci Národní digitální knihovny:

parametr	význam
Cblk="{64,64}"	velikost bloku - zvednuto z 32 na 64
Corder="RPCL"	progress order, rozlišení - pozice - komponenta - vrstva. Ideální pořadí pro image servery.
Stiles="{16384,16384}"	použití dlaždic se budeme ve většině případů snažit vyhnout
ORGgen_plt="yes" ORGtparts="R"	značky pro lokalizaci regionů
Cprecincts="{256,256},{256,256},{128,128}"	velikost regionů
Creversible=yes	beztrátová komprese, filtr (9,7) a RCT (reversible colour transformation)
-rate -, 4, 2.34, 1.36, 0.797, 0.466, 0.272, 0.159, 0.0929, 0.0543, 0.0317, 0.0185	kompresní poměr pro jednotlivé vrstvy, - značí beztrátovou kompresi.
Clayers=12	počet vrstev kvality
Clevels=6	počet úrovní rozlišení pro rychlé generování náhledů, bude se lišit v závislosti na velikosti obrázků tak, aby nejmenší náhled byl v rozmezí 128x128 až 256x256.
Cmodes={BYPASS}	urychlí kompresi a dekompresi

Moravská zemská knihovna tak již v současnosti vytváří digitalizované mapy ve formátu jpeg2000 s těmito parametry.

A.6 Vybrané relevantní články a odkazy

Simon, Korb, Sadilek and Schmidt (2009) Collaborative Map Annotation in the Context of Historical GIS, Proceedings of the Workshop on Geospatial Computing for the Arts, Humanities and Cultural Heritage, Oxford, UK. <http://userver.ftw.at/~simon/files/SW095001.pdf>

Grosso, Bouju, Mustière (2009): Data Integration GeoService: A First Proposed Approach Using Historical Geographic Data. W2GIS 2009: 103-119, <http://goo.gl/H8BRe>

Simon, Korb, Sadilek, Baldauf (2010) Explorative User Interfaces for Browsing Historical Maps on the Web, Proceedings of the 5th International Workshop on Digital Approaches in Cartographic Heritage, February 22-24 2010, Vienna, Austria. <http://userver.ftw.at/~simon/files/cartoheritage-paper29.pdf>

Simon, Sadilek, Korb, Baldauf, Haslhofer (2010) Tag Clouds and Old Maps: Annotations as Linked Spatiotemporal Data in the Cultural Heritage Domain. In Workshop On Linked Spatiotemporal Data 2010, Zurich, Switzerland, 2010.

B Analytická část

I v letošním roce byla největší část práce věnována analýze již dokončených nástrojů a programátorské práci. Nemalý čas strávili členové týmu i přednáškovou činností a poskytováním konzultací.

B.1 Manipulace s rastrovými daty

B.1.1 MapTiler

V letošním roce byla dokončena lokalizace aplikace MapTiler do češtiny prostřednictvím online nástroje Launchpad (<https://translations.launchpad.net/maptiler/>). Mimo češtiny přibyla prostřednictvím tohoto nástroje díky dobrovolníkům i italština, rumunština, katalánština, téměř dokončena je i španělština. Nová verze MapTileru bude pravděpodobně publikována v příštím roce.

B.1.2 Online prohlížení dlaždic - OpenLayers

Zoomify layer, vyvinutý v loňském roce, byl po zapracování všech připomínek ze strany vývojářů OpenLayers akceptován do distribuce OpenLayers. Toto je sice krátké konstatování, ale skrývá se za ním velké množství práce, která se mnohonásobně zúročí tím, že nebude nutné patch pro Zoomify udržovat vlastními silami a bude přitom zajištěna podpora pro Zoomify v nových verzích OpenLayers. Navíc díky tomuto kroku je už nyní využívána naše práce v jiných výzkumných projektech a online aplikacích - je to tedy jeden z velmi významných výsledků projektu OldMapsOnline.org.

B.1.3 Scaling Zoomify Tileset

V současnosti probíhají práce na vývoji rozšíření funkcionality Scaling TileSet pro Google Maps verze 3 s možností zobrazit naskenované mapy v podobě jednoduchých, korektně zvětšených a zrotovaných dlaždic na základě námi dodaného algoritmu pro výpočet pozice a pootočení každé dlaždice a kódu pro získání jednotlivých Zoomify dlaždic. Funkce pro rotaci bude využívat funkci Closure library goog.graphics, která na pozadí využívá SVG, Canvas nebo VML podle toho, ve kterém prohlížeči je spuštěna. Je založena na <http://closure-library.googlecode.com/svn/trunk/closure/goog/demos/graphics/advancedcoordinates.html>

Jádrem práce je změna logiky pro zobrazování, nahrávání a odstraňování dlaždic které jsou viditelné v okně mapy. Nové dlaždice jsou rotovány, nahrány, neproporcionálně zvětšeny nebo zmenšeny s úrovněmi zvětšení neodpovídající pyramidě sférického Mercator zobrazení.

Výsledek bude velmi podobný tomu co dělá demo na <http://polymaps.org/ex/transform.html>, ale prohlížení bude fungovat i na Internet Exploreru, díky knihovně Closure. Uvedené demo sice podporuje plnou afinní transformaci se 6 operacemi, v našem případě bude ale operací jen 5 (nebudeme podporovat zešíkmení), abychom zajistili podporu ve všech prohlížečích.

Vývojem této funkcionality byl pověřen fin Esa Ilmari, Googlem certifikovaný vývojář, který má s podobnou prací rozsáhlé zkušenosti, jak dokládá například přehled jeho práce na <http://sites.google.com/site/esailmari/>. Zadaná práce bude dokončena ještě v letošním roce. Tento projekt umožní zmenšit zátěž našeho serveru, který vypočítává transformace map georeferencovaných v nástroji Georeferencer.org a poskutekuje ji formou standardu OGC WMS. Část této funkcionality tak přesouváme na stranu klienta - do webového prohlížeče.

B.1.4 Dynamické generování dlaždic – image servery

Jak již bylo zmíněno, v letošním roce pokračuje vývoj image serveru IIPImage, přičemž jeho vývoj vychází zejména z analýz prováděných v rámci tohoto projektu. Pro výkon image serveru je klíčová

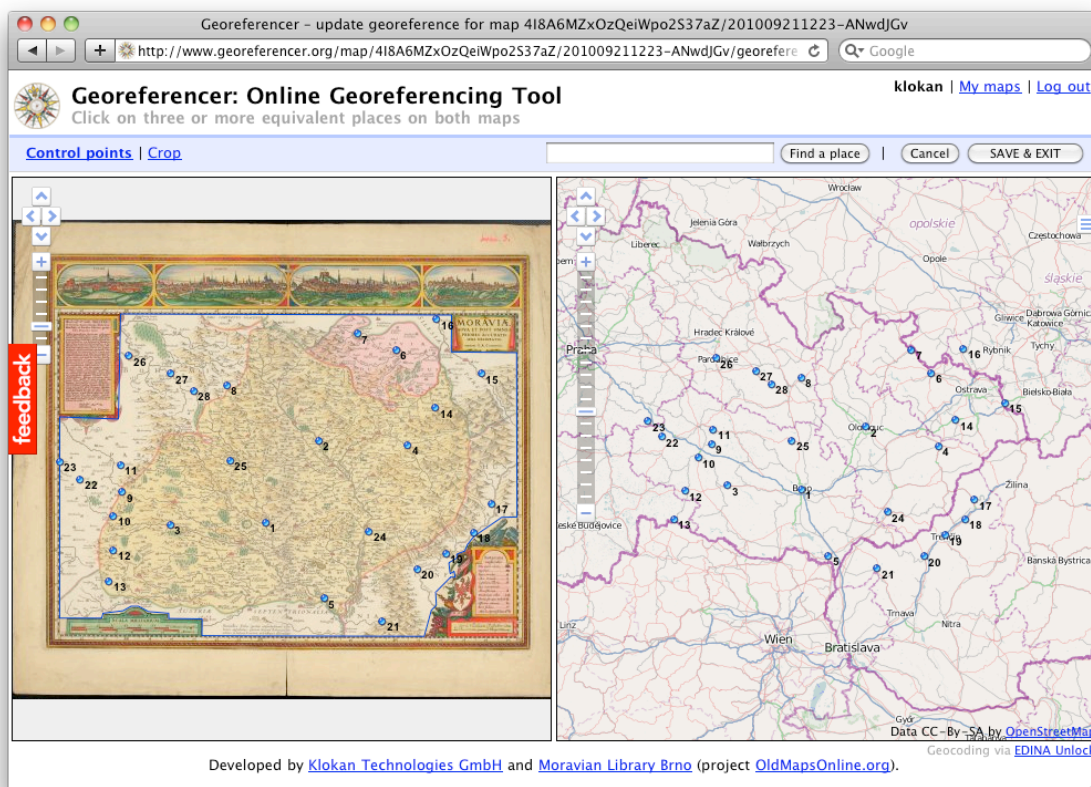
implementace Memcached, která přináší výrazné zrychlení image serveru a ve spojení s http serverem nginx má potenciál být velmi rychlým image serverem zejména pro často požadované dlaždice. V minulé zprávě byla taková aplikace již zmíněna, jde o obdobu aplikace <http://Zoom.it> Microsoft Research, která funguje jako image cache - uživatel nahraje na server rozměrný obrázek a služba jej zpřístupní v podobě dlaždic, v případě Zoom.it v podobě dlaždic DeepZoom. Taková služba je však náročná na diskový prostor, výkon serveru propustnost internetové přípojky, jak ukazuje její testovací interní provoz v MZK a proto se snažíme přesvědčit k jejímu zprovoznění Europeanu, které by taková aplikace mohla pomoci vyřešit problém s anotacemi obrázků (ty se nedají dost dobře dělat dokud European sklízí pouze metadata).

V souvislosti s již zmíněnou integrací imageserveru do nové verze digitální knihovny Kramerius se ukázalo jako nezbytné vylepšit podporu ukládání a zobrazování map v Krameriu. Jednak jde o zavedení kartografických dokumentů jako samostatného druhu dokumentů a dále zajištění permanentnosti URL, prostřednictvím kterého je mapa zpřístupňována - MZK totiž zatím staré mapy zpřístupňuje prostřednictvím image serveru umístěného na URL imageserver.mzk.cz mimo Krameria a vzhledem k tomu, že Georeferencer pracuje s URL jako identifikátorem georeferencované mapy, je nutné zajistit neměnnost těchto identifikátorů po integraci map do Krameria. Tím bude zároveň položen základ neměnnosti těchto URL obrázků v Krameriu obecně.

B.2 Kartografické transformace

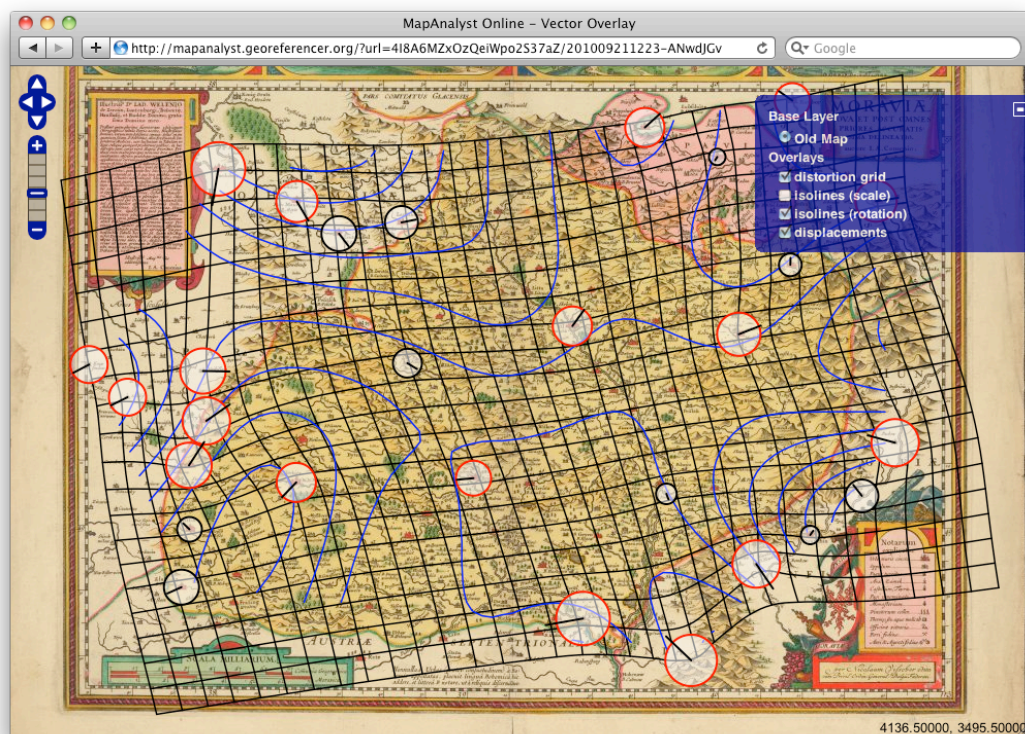
B.2.1 GeoReferencer.org

Na adrese <http://www.georeferencer.org> je zprovozněna betaverze georeferencovacího nástroje. Ta má oproti první verzi zcela přepracované georeferencovací rozhraní, kde se přešlo od filozofie map posouváných pod pevným kurzorem k filozofii dvou nezávislých map s vyznačováním bodů na libovolném místě zobrazené části mapy, což práci s nástrojem udělalo intuitivnější a rychlejší.



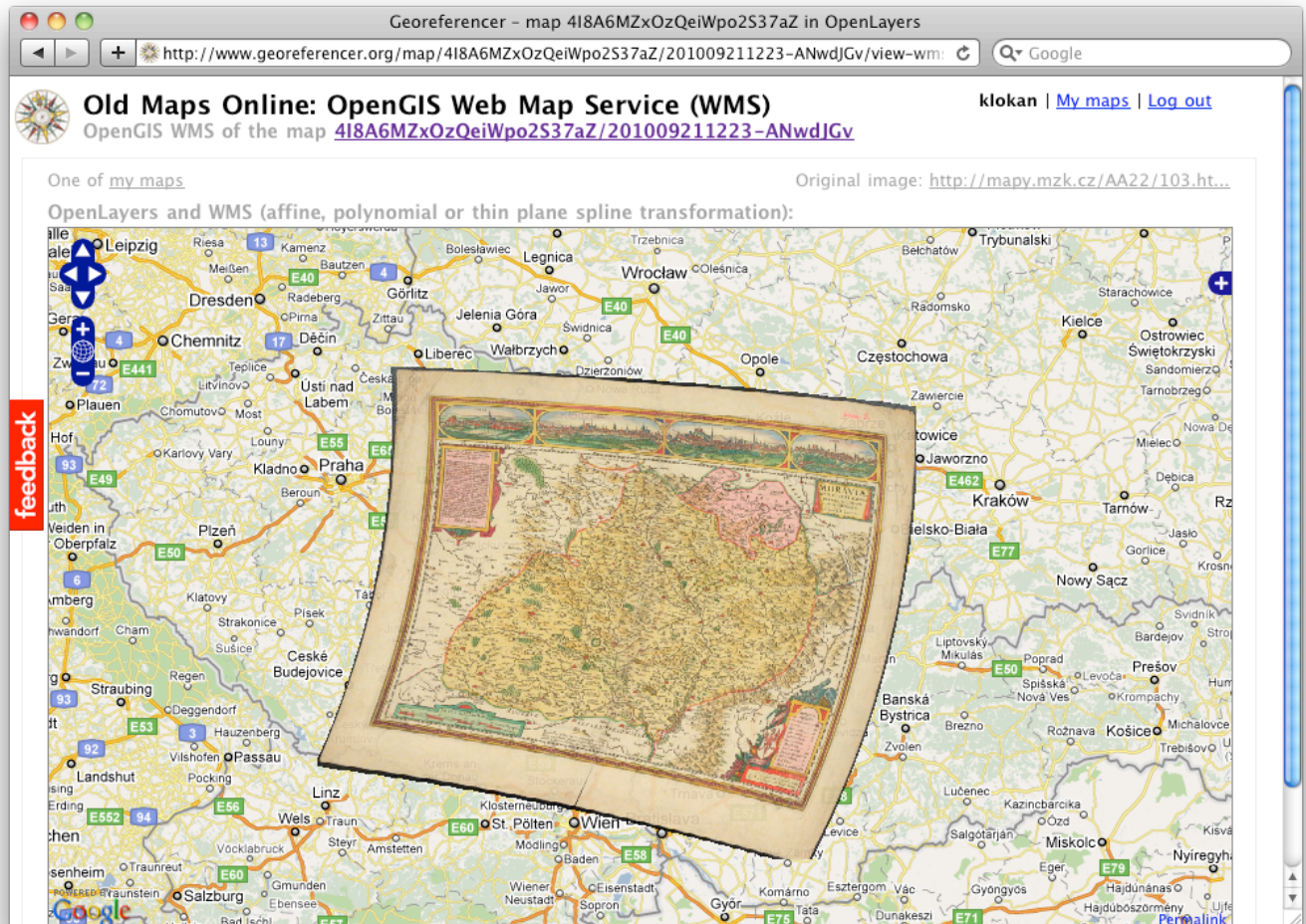
V rámci vývoje betaverze Georeferencer došlo k mnoha dílčím krokům:

- Optimalizace provozu na Google App Engine, kam se ukládají i vytvořená metadata a uživatelské účty (ve špičce Google využívá paralelně až 8 serverů na obsluhu online služby Georeferencer).
- Zavedení možnosti zadat ořezovou křivku mapy (část obrázku, skutečně zobrazující mapu)
- Zavedení možnosti editace bodů - jejich přesun, mazání
- Podpora systematického georeferencování mapové sbírky instituce - instituce dodá seznam URL a zpět dostane formou tabulky (Google spreadsheet, Microsoft Excel) výsledky georeferencování přes online rozhraní. K vizualizaci seznamu map bylo využito aplikace CoolIris a převodu seznamu odkazů do dvou RSS feedů: jeden pro mapy které daná instituce potřebuje zgeoreferencovat a druhý, s mapami již georeferencovanými (ale je možno je dále vylepšovat). Příkladem takového rozhraní je
 - <http://www.georeferencer.org/maps/domain/imageserver.mzk.cz/cooliris> s RSS feedy
 - <http://www.georeferencer.org/maps/domain/imageserver.mzk.cz/ungeoreferenced.rss>
 - <http://www.georeferencer.org/maps/domain/imageserver.mzk.cz/georeferenced.rss>
- Přechod od přihlašování účtem od Googlu k přihlašování pomocí vlastního jména a hesla - přihlašování Google účtem některé uživatele mátló.
- Tlačítko „Georeferencer“ (tzv. bookmarklet), které je možné přetáhnout na lištu prohlížeče, bylo výrazným způsobem vylepšeno. Jeho kód je umístěna na našem serveru, takže je možné jeho funkcionalitu průběžně aktualizovat aniž by uživatelé byli nuceni jej znovu instalovat po změně.
- Přímá integrace nástroje MapAnalyst Online pro vizualizaci kartometrické přesnosti mapy na základě bodů zadaných ve webovém rozhraní Georeferencer.



- Integrace UMN MapServeru linkujícího na vzdálené dlaždice Zoomify a zajišťujícího přímé warpování dlaždic formou WMS proxy. Toto však výrazně zatěžuje server a ubírá tak na rychlosti

a škalogatelnosti celého systému Georeferencer. V planu je nasadit před MapServer WMS ještě TileCache a v budoucnu také implementovat alternativní systém (s pomocí HTML5, SVG, VML, Canvas, či WebGL), který by přesunul zpracování obrazu na stranu klienta, pokud má návštěvník webový prohlížeč podporující příslušné technologie. Server by tak byl méně zatěžován.



- Georeferencer podporuje nyní více druhů dlaždic. Je tam možné georeferencovat kromě formátu Zoomify i online obrazy publikované pomocí nástrojů jako ContentDM nebo produktu firmy ExLibris (např. DigiTool), a několika dalších způsobů zpřístupnění obrazů ve vysokém rozlišení.

B.2.2 Warpování v browseru

Pro warpování map v Browseru nadále počítáme s využitím tagu HTML5 canvas, s tím rozdílem, že rendering mapy nebude zajišťovat prostý výpočet v javascriptu, ale WebGL. Protože jde o poměrně specializovanou funkcionalitu, vyžadující specifické programátorské znalosti, byla tato práce zadána programátorovi, který již má zkušenosti s programováním v OpenGL. Výsledky jeho práce lze očekávat v pololetí příštího roku.

B.3 Geografické vyhledávání map

V závěru roku 2009 byla provedena srovnávací analýza dostupných systémů pro vyhledávání map. Tato analýza byla zaměřena zejména na nalezení vhodného nástroje pro geografické prohledávání rozsáhlých databází mapových sbírek. Zatímco běžné knihovní katalogy takovou funkčnost vůbec nepodporují, protože jsou založené výhradně na textovém vyhledávání, existuje několik specializovaných aplikací, které podobnou funkčnost nabízejí. Jedna z nich (TimeMap) byl v MZK v minulosti testován a je nasazen na adrese <http://timemap.mzk.cz>. Toto řešení lze považovat za funkční prototyp, který po formální

stránce plně splňuje cíle projektu Staré mapy online.

Pokud je však cílem vytvořit takové řešení, na kterém lze stavět i v budoucnu, nelze pominout některé nedostatky aplikace TimeMap. Největším nedostatkem této aplikace je to, že se její vývoj před několika lety zastavil a lze ji tak dnes považovat za mrtvou - sice funguje, ale její další rozvoj by si vyžádal její úplné přepracování a vázal by neúměrně vysoké prostředky (pravděpodobně miliony Kč a mnoho času). Další problém je koncepční: TimeMap je javová aplikace, která počítá s tím, že běží v prohlížeči jako java applet (což je samo o sobě již dnes překonané řešení) a navíc počítá s tím, že celý prohledávaný dataset je stažen na lokální počítač a tam prohledáván. To je velmi neefektivní řešení zejména při rostoucím počtu prohledávaných dokumentů, protože celý vyhledávací proces pak závisí na rychlosti daného java appletu a velké počty dokumentů hledání výrazně zpomalují. Dalším problémem pak je způsob zobrazení výsledků vyhledávání. Výsledky jsou zobrazeny v podobě tabulky, nejsou však nijak řazeny. Vyhledávání navíc nezohledňuje měřítko hledaných map.

Z těchto důvodů jsme se pokusili najít takové řešení, které by nemělo výše uvedené problémy. Hledané řešení by mělo splňovat přibližně tyto parametry:

- pracuje s bibliografickými záznamy které obsahují přinejmenším souřadnice hraničního obdélníka (tzv. bounding box) každé mapy. Díky tomu by bylo možné pracovat s MARC21 záznamy obsahujícími tyto údaje v poli 034.
- intuitivní uživatelské rozhraní, v optimálním případě využívající některou z populárních online mapových služeb coby podkladovou mapu světa, nad níž bude prováděno vyhledávání.
- podpora zadání časového intervalu na časové ose
- podpora omezení vyhledávání zadáním fulltextového řetězce
- podpora omezení vyhledávání podle měřítka mapy
- podpora rankingu - řazení výsledků vyhledávání podle relevance
- škálovatelnost: okamžitá reakce na změnu parametrů vyhledávání nezávisle na velikosti prohledávané databáze
- integrovatelnost do webů institucí, propojitelnost s nástroji vyvíjenými v projektu Staré mapy online.

V rámci analýzy jsme mimo aplikace TimeMap, jejíž nedostatky jsou popsány výše, testovali následující nástroje:

1) GeoNetwork (<http://geonetwork-opensource.org/>) je open source software, používaný pro správu katalogu geodat a geoslužeb. Podporuje mnoho standardů pro import, konverzi a export metadat. Má i možnosti prohledávání indexovaných databází metadat. Toto prohledávání je postaveno na komponentě Apache Lucene, jejíž primární funkcionalitou je textové hledání. Geografický ranking není vůbec implementován, jen základní omezení na geografickou oblast ve které musí být metadata obsažena. Třídění je řešeno na úrovni fulltextu. Projekt je poměrně komplexní a hodí se především pro nasazení v GIS prostředí. Úpravy pro potřeby knihoven (které již většinou mají nasazen jiný existující knihovní systém používaný primárně pro katalogizaci) by byly poměrně komplikované. Navíc problém geografického třídění v projektu není řešen.

2) Lucene with GeoTemporal Extensions (LGTE)

(<http://code.google.com/p/digmap/wiki/LuceneGeoTemporal>) vyvinutý v rámci evropského projektu DigMap.eu přidává do Lucene enginu podporu pro geografické a časové hledání. Primárním problémem je ale že dokumenty mají záznam formou geografického bodu, nikoliv oblasti kterou pokrývají. Geografický ranking je řešený vzdáleností takového bodu od diagonály dotazovacího okna, což rozhodně není pro prohledávání většího katalogu map ideální. LGTE se ale hodí skvěle k jiným účelům, kde reprezentace dokumentu formou bodu vyhovuje - především je používán ve spojení IR systémy geograficky indexující textové dokumenty. Modifikace Lucene pro geografické třídění narážejí také na problémy s rychlostí odpovědi při větším množství indexovaných dokumentů a komplikovanějším dotazu.

3) PostGIS (<http://postgis.refractory.net/>) a jiné tradiční GIS databázové systémy - poskytují možnost rychlé a efektivní indexace velmi rozsáhlých geografických dat s možností prohledávat pomocí SQL jazyka díky podpoře OpenGIS "Simple Features Specification for SQL". Tradiční algoritmus indexace v prostorových databázích používá R-strom. Ten vrátí pro geografický dotaz všechny záznamy ve zvolené oblasti. Sám o sobě ale neřeší třídění - to je plně na aplikaci.

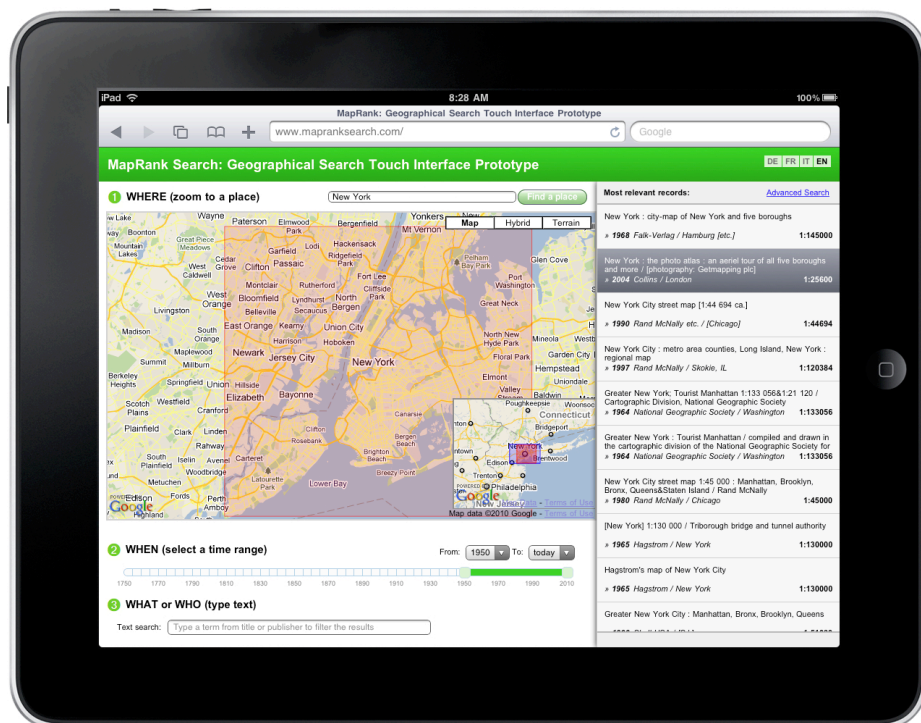
4) Alexandria Digital Library Project (<http://www.alexandria.ucsb.edu/>) využívá PostGIS pro indexaci a prohledávání geometadat v prostředí digitálních knihoven. Přidává navíc podporu pro základní formu geografického třídění pomocí algoritmu Hausdorff distance (http://en.wikipedia.org/wiki/Hausdorff_distance). Finální pořadí výsledků se vypočítává z množiny všech kandidátů vrácených pomocí indexace R-stromem v okamžiku dotazu uživatele - to znamená, že pro větší množství indexovaných dokumentů je pak vyhledávač velmi pomalý pro oblasti, které mají příliš mnoho překryvů. Například při hledání mapy Prahy jsou nalezeny všechny mapy, které pokrývají oblast Prahy, včetně map Evropy a světa. S indexováním pomocí R-stromu se pak pořadí podle relevance počítá pro všechny tyto mapy na straně aplikace - také to tedy není optimální řešení pro prohledávání databáze velkého množství map. Ranking pomocí Hausdorff distance navíc nevrací ideální přirozené pořadí výsledků.

5) Prohledávání metadat publikovaných na webu formou vhodnou pro indexaci pomocí Google Geo Search API (<http://googlemapsapi.blogspot.com/2008/05/geo-search-20-data-in-data-out.html>) nebo Google Fusion Tables (<http://googlegeodevelopers.blogspot.com/2010/11/search-your-geo-data-using-spatial.html>). Testovali jsme i tuto formu prohledávání, ale problém je opět chybějícího třídění výsledků s jinou interní reprezentací než formou geo bodu. Navíc je nutné vlastní metadata plně předat společnosti Google.

Při exportu geo-metadat formou vytvoření KML a publikování souboru SiteMap.xml připojující KML metadata ke každé georeferencované mapě opět hledání selhává pro staré mapy, protože Google i další vyhledávače indexují jen textová data vztahující se k určitému místu a neřeší ani ranking ani bounding boxy.

Provedená analýza tedy ukázala, že nástroje pro vyhledávání map, využívající kombinaci, fulltextu, geografických souřadnic a času, pokud vůbec existují, postrádají podporu optimálního řazení výsledků (ranking) tak, aby uživatel obdržel v průběhu vyhledávání vždy ty nejrelevantnější výsledky. Přestože jsme předpokládali že nalezneme takový systém, ukázalo se že podobná technologie není v současnosti nikde dostupná ani jako open source ani jako komerční aplikace. Implementace vlastního rankingu a systému indexace dokumentů jde však nad rámec projektu Staré mapy online, ve kterém se počítalo s využitím TimeMap nebo jeho alternativy, nejvýše s mírnou úpravou.

Proto byl v loňském roce připraven projekt do programu VISK, který tento problém řeší zakoupením celonárodní licence systému MapRank Search, který byl původně vyvinut pro potřeby prohledávání souborného katalogu Švýcarských knihoven, a je již nasazen v ostrém provozu na portálu Kartenportal.ch (<http://www.kartenportal.ch/>). V MZK je systém v době psaní tohoto textu zprovozněn na adrese <http://mapranksearch.mzk.cz>.



Algoritmus MapRank použitý pro třídění poskytuje:

Optimální pořadí map ve výsledcích hledání: díky specifické implementaci indexace geometadat je možno využít kvalitnějších metod hodnocení pořadí dokumentů. Přirozené řazení podle prostorové podobnosti je realizováno na základě výpočtu vzdáleností a ploch na zemském povrchu. Rychlost: i přes kvalitu třídění výsledků vyhledávání jsou dotazy systémem zodpovězeny ve zlomku sekundy.

Systém MapRank Search aplikuje výpočet prostorové podobnosti oblasti specifikované uživatelem (vyhledávací „dotaz“) a oblasti, kterou pokrývá každý jednotlivý indexovaný dokument. Takové numerické vyjádření podobnosti coby "vzdálenosti" dvou geografických oblastí je pak použito pro řazení výsledků v kombinaci s ostatními filtry, jako fulltext nebo časové období. Řazení dokumentů je velmi přirozené, je to to, co člověk očekává od pojmu "geografická podobnost", ve smyslu podobnosti tvaru, velikosti a místa i prostorového překryvu.

B.4 Navazující projekt

Vzhledem k tomu, že pro rok 2011 je již naplánován jen velmi malý objem financí, projekt Staré mapy online fakticky v roce 2010 končí. V průběhu jeho řešení se však objevilo značné množství dosud nevyřešených problémů, které si zasluhují další výzkum a vývoj. Proto byl ve spolupráci s Univerzitou Karlovou a Masarykovou univerzitou připraven pokračující projekt do programu NAKI.

C Návrhová část

Ve třetím roce řešení projektu bylo již dosaženo všech výsledků potřebných pro úspěch projektu. Ještě v listopadu 2010 tak bude dalším knihovnám nabídnut postup, jak si mohou s využitím Georeferenceru zpracovat vlastní sbírky digitalizovaných map. Vzhledem k rozpočtu projektu na rok 2011 bude většina činností pravděpodobně ukončena v průběhu prvních měsíců roku tak, aby v případě schválení navazujícího projektu podaného do programu NAKI bylo možné plynule navázat.

Petr Žabička ještě v letošním roce seznámí s výsledky projektu účastníky 11. konference Archivy, knihovny, muzea v digitálním světě (<http://skip.nkp.cz/akcArch10.htm>), konané ve dnech 1.-2.12.2008 v Praze. Prezentace budou pokračovat na relevantních fórech i v příštím roce.

Podrobný přehled vývojových prací je včetně specifikace priorit (metoda MuSCoW) zveřejněn na adrese <http://sites.oldmapsonline.org/todo>. Z dokumentu je zřejmé, že již byly dokončeny všechny nezbytně potřebné komponenty a s jednou výjimkou, která je ve stavu alfaverze i komponenty s prioritou "should have".

Výsledky uplatnitelné v RIV

Článek v českém recenzovaném periodiku (Jrec):

- Žabicka, Petr. Volně dostupné nástroje z produkce projektu Staré mapy online. Knihovna [online]. 2010, roč. 21, č. 1. Preprint dostupný z WWW: <http://www.mzk.cz/projekty/oldmapsonline/Knihovna-clanek-staremapy.pdf>. ISSN 1801-3252.

Uplatnění dalších výsledků bude záviset na skutečné podobě pokynů k poskytování výsledků pro RIV v průběhu jara 2011. Vzhledem k tomu, že v průběhu jara 2010 se tato pravidla čtyřikrát změnila, přičemž v některých případech šlo o změny mající dopad na uznatelnost výsledků, lze předpokládat podobný vývoj i v roce 2011 a nelze tak predikovat, které výsledky bude možné uplatnit.

Ověřená technologie/poloprovoz:

- Georeferencer - linka pro georeferencování digitalizovaných starých map
- MapAnalyst online - ověřená technologie pro vizualizaci deformací starých map

Výsledky pravděpodobně neuplatnitelné v RIV

Článek ve sborníku:

- P. Pridal (2010) OldMapsOnline.org: IIPImage, JPEG2000 and Georeferencer.org, Proceedings of the 5th International Workshop on Digital Approaches in Cartographic Heritage, February 22-24 2010, Vienna, Austria.

D

E Stručné resumé

Ve třetím roce řešení projektu jsme dokončili vývoj všech komponent důležitých pro úspěch projektu. V současné době probíhají dokončovací práce, spočívající ve vylepšování uživatelského rozhraní, lokalizaci, zlepšování vzájemného propojení komponent, vylepšování funkčnosti a v neposlední řadě hledání způsobů zajištění dlouhodobé udržitelnosti vyvinuté infrastruktury. Vzhledem k povaze projektu nespočívá udržitelnost v prostém udržení systémů v provozu, ale zejména v jejich dalším rozvoji tak, aby reflektovaly nejnovější trendy. Proto jsme ve spolupráci s Univerzitou Karlovou a Masarykovou univerzitou připravili a podali nový projekt do programu NAKI. Vzhledem k tomu, že vyvinutá infrastruktura je celosvětově unikátní, je naší snahou uplatnit výsledky projektu na mezinárodní úrovni, například v rámci portálu Europeana.

Nejdůležitějším počinem letošního roku je bezesporu dokončení betaverze nástroje pro georeferencování, který disponuje oproti původní verzi značně přepracovaným uživatelským rozhraním a podporuje také širší škálu různých image serverů a dlaždicových struktur.

Dalším důležitým milníkem je dokončení nástroje MapAnalyst Online a jeho přímé napojení na GeoReferencer. Díky tomu je nyní možné jednoduše vizualizovat uživateli během georeferencování vložené údaje.

Již nyní je tak možné nabídnout paměťovým institucím zpřístupňujícím digitalizované mapy možnost je jednoduchým způsobem začít georeferencovat. Praktické testování vyvinutých nástrojů v ostrém provozu nyní probíhá.

F Přílohy

F.1 Zprávy ze služebních cest

F.1.1 Petr Přidal, GIR Workshop, Curych, 18.2.2010

Mluvil jsem pomerne dost s Bruno Martinez, a tahal z neho informace o tom co se deje kolem Borbinhy a ve Spanelsku. Prof. Borbinha (chair of latest ECDL) planuje novy EU projekt s nejakymi institucemi z UK: http://www.york.ac.uk/res/isthmus/ISTHMUS_Site/ISTHMUS_Project.html

Jorge Machado, clovek co prakticky naprogramoval LGTE: <http://en.wikipedia.org/wiki/LGTE> by eventulne byl ochoten spolupracovat na dalsim rozvoji tohoto nastroje.

Mam novy kontakt do Google Zurich, muzu toho cloveka pouzit k prosazovani opravy bugu ci requestu pro novou funkcionalitu v produktech Google. On to pushne dal ve firme, nebo to preposle relevantnim lidem. Je to engineer zabývající se hledáním na Google Maps, typu "pizza New York" ;-).

Zajimavy kontakt tykající se rankingu by mohl byt clovek vedouci Yahoo Research v Barcelone: http://research.yahoo.com/Hugo_Zaragoza.

Nejaci lide z nemecka pry pracuji na Spatial Similarity, vypada to na tyhle: http://ifgi.uni-muenster.de/~janowicz/wp-content/uploads/giscience2008_Similarity-based_Information_Retrieval_and_its_Role_within_Spatial_Data_Infrastructures.pdf

V Edinburku v ramci projektu EDINA vyvinuli webovou sluzbu podobnou Yahoo! PlaceMaker, postavenou na Geonames (a tedy fungujici i s vlastnim Gazetterem, ktery muzeme doplnovat a upravovat), ktera je primo pouzitelna pro hledani toponym v metadatech - tedy ve sluzbe postavene na mem prototypu <http://geoparser.appspot.com/>. Pro pouziti s ceskymi metadaty by bylo vhodne predzpracovat text do formy lemmat (zakladniho tvaru slov) externim nastrojem, jako je AJKA z FI MUNI, tedy "Mapa Brna" - > "Mapa Brno", jinak je sluzba primo pouzitelna. Adresa teto webove sluzby je: <http://unlock.edina.ac.uk/text.html>

Prezentace Bruno Martineze se zabyva kombinaci rankingu do jednoho vysledku, coz by se mohlo hodit pozdeji.

Web conference: <http://www.geo.unizh.ch/~rsp/gir10/>

F.1.2 Petr Přidal, Petr Žabička – 5th International Workshop on Digital Approaches in Cartographic Heritage, Vídeň, 22.-24.2.2010

Na konferenci jsme meli prednasku s nazvem "OldMapsOnline.org: IIPImage, JPEG2000 and Georeferencer.org", na ktere jsme prestavili verejnosti vysledky z grantu. Demo Georeferencer.org z prednasky je nyní online na: <http://www.vimeo.com/9709771>

Behem konference jsme se potkali s lidmi, ktere by potencialne mohli byt partnery v nasledujicim evropskem projektu, predevsim pak:

- Austrian Institute of Technology - kontakt pro organizaci grantu je Joachim Korb <joachim.korb@ait.ac.at> a jako vedouci vyvojar pak Simon Rainer <simon.rainer@ait.ac.at>. Jsou momentalne aktivni na projektu Europeana Connect. Chteji aby som v zari predvedli nase nastroje na konferenci tykající se Europeana Connect.

- Institute of Cartography - ETH Zurich - kontakt je Bernhard Jenny <jenny@karto.baug.ethz.ch>, po organizační stránce asi spis jeho šef (který ve Vídni nebyl) - prof. Lorenz Hurni.

Další dva zajímavé kontakty, které jsou spíše potenciálními uživateli technologií na kterých pracujeme jsou:

- ICC Barcelona - kontakt: Rafael Roset (s tím jsme se viděli už v Bratislavě na ELAGu)
- National Archive of Netherlands: kontakt: Jeroen van Luin <jeroen.van.luin@nationaalarchief.nl>

Jako součást konference jsme také zorganizovali brain-storming session, s lidmi které naše nástroje zajímají. Ta byla velmi přínosná pro nás jako vyvojáře Georeferencer.org. Zúčastnilo se asi 12 lidí.

Diskutovali jsme budoucí funkcionality, potenciální problémy a jejich řešení, metadata a algoritmy týkající se georeference a mapových projekcí. Také jakým způsobem navrhnout nástroje tak aby přilákaly nové návštěvníky na web knihovny.

Dne 25.2. bylo v MZK uspořádáno setkání s potenciálními partnery projektu, a to s Geografickým ústavem MU. Dále bylo koordinační setkání všech pracovníků MZK pracujících na projektu.

F.1.3 Petr Přidal, coding session v MZK, Brno, 15.-20.8.2010

Neděle večer přelet z Zurichu do Prahy, transport do Brna. Pondělí až čtvrtek školení a kódování společně s lidmi z MZK. Pátek návrat přes Prahu do Zurichu a následně do bydliště v Baar, CH.

Aktivní workshop s lidmi v MZK, předání zkušeností s novými technologiemi jako: Google Closure Library, App Engine, Amazon EC2. Týmový vývoj komponent pro Georeferencer.org. Jednání ohledně dalšího vývoje tohoto nástroje a určení priorit pro novou funkcionality.

Účastníci workshopu si vyzkoušeli Closure Library. Diskuze nad technologiemi, dohoda na výměně dat, identifikátorech, permanentních URL a formě komunikace mezi komponentami – vše v týmu. Praktickým výsledkem za tento týden jsou:

- Nové georeferencovací rozhraní na Georeferencer
- Praktické použití Amazon EC2, úpravy kódu Georeferencer služby
- MapAnalyst integrace přes JSONP API, nová prohlížečka na mapanalyst.georeferencer.org
- Port CoolIris RSS prohlížečky do AJAXu
- DublinCore metadata editor v JavaScriptu
- MZK ImageServer s permanentními URL: (<http://imageserver.mzk.cz/MZK03/001/042/265/2619266100.jp2>)

F.1.4 Petr Přidal, Konference FOSS4G, Madrid, 6.-13.9.2010

Příjezd v pondělí večer, úterý zahájení, konference až do čtvrtka, v pátek coding-session. V pondělí setkání s lidmi z ICC (Rafael Roset a spolupracovníci), večer pak návrat.

Prezentace výsledků projektu MZK: „OldMapsOnline.org“, setkání s lidmi z ICC.cat (Spain), a během konference pak s vývojáři z NYPL.org (USA), a s lidmi z EDINA Unlock (UK). Diskuze vývoje Georeferencer.org, plánu do budoucna a vzájemná synchronizace vývoje a výzkumných aktivit.

Prezentace je online: <http://www.vimeo.com/14967243>, byla velmi úspěšná (podle ohlasů). Letáky k projektu OldMapsOnline.org, které jsem dovezl, byly rozebrány ihned po prezentaci. ICC dodalo připomínky k Georeferencer.org, měli by poslat technické informace ke ContentDM systému, který chceme v Georeferencer.org podporovat. Diskuze s lidmi stojícími za NYPL MapRectifier (Tim Waters, Schuyler Erle) byla naprosto klíčová pro další směřování vývoje Georeferencer. Pracují na podobném projektu – viděl jsem prototypy a diskutoval další rozvoj jejich služby. Jednání s lidmi s EDINA Unlock, JISC (UK), jejichž API používáme, umožnilo dodat připomínky pro jejich službu. Díky tomu plánují rozšíření jejich služby tak, aby nám více vyhovovala.

<http://2010.foss4g.org/program.php>

Podrobnější zpráva je publikována online na <http://duha.mzk.cz/blog/vysledky-rd-projektu-mzk-stare->

F.1.5 Petr Přidal, Odborný seminář Digitalizace mapových sbírek a archivů, Praha, 21.-23.10.2010

Cílem cesty na odborný seminář Digitalizace mapových sbírek a archivů bylo vyslechnutí nových výsledků týkajících se vývoje nástrojů pro publikování map a také postup digitalizace map v jiných institucích. Povedlo se částečně i oznámení postupných výsledků vývoje v MZK. Důležité bylo osobní setkání především s lidmi z ČUZK, ČVUT, AV a Karlovy univerzity a diskuse budoucí spolupráce týkající se starých map. Důležitá byla i následná osobní diskuse s lidmi z ČUZK (Meissner), VuGTK (Talich), ČVUT (Veverka, Cajthaml) a AV (Chodějovská).

- CUZK: <http://archivnimapy.cuzk.cz/> - zajímavý je kompletní soubor map pro ČR "S-1952" - možný kandidát pro českou verzi out-of-copyright mapy území státu
- Národní archiv Praha: "JANUS" vs. "Bach": inventarizační software používaný v archivech. Poznámka: "Žádoucí je natáčení map v prohlizecce!". <http://www.badatelna.cz/> - používá Zoomify!
- BTW Proč nikdo v archivech nepoužívá standardy jako DublinCore a OAI-PMH společně se zdarma softwarem jako je např. open-source <http://www.omeka.org/> nebo i zdarma online služba <http://www.omeka.net/> ?
- Národní zemědělské muzeum/Kartografické centrum Velké Opatovice. Připravují publikování nějakých naskenovaných map na Internet. (Paulus)
- Chodějovská: Historický ústav AV ČR - digitalizace společně s VuGTK: Zoomify na <http://mapy.vugtk.cz/>
- Historicky atlas měst (online) společně s ČVUT: <http://maps.fsv.cvut.cz/praha/vinohrady> na VuGTK volné WMS pro: II. vojenské mapování i III. vojenské mapování, pozemkový katastr
- Česká geologická služba: <http://www.geology.cz/>
- UK: Přírodovědecká fakulta (Novotná): 130 naskenovaných map v Digitool, 400 záznamů v Alephu.
- ČUZK: Připravovaná aplikace p.mapper na ČUZK: <http://archivnimapy.cuzk.cz/mapy/>
- Slezský archiv v Opavě (Peterka): ICARUS: <http://www.icar-us.eu/> : Mezinárodní centrum pro archivní výzkum (ve Vídni). Prezident: "Dr. Thomas Aigner". <http://www.icar-us.eu/> . Monasterium (100 000 listin).
- Projekt Svědectví map (digitalizace starých lesnických map): <http://www.nzm.cz/prodej-publikaci-predmetu/>

F.1.6 Petr Přidal, Petr Žabička, Place, Cultural Heritage and the InternetEuropeanaConnect - Cross-Project Networking Workshop,Vídeň, 4.-5.11.2010

Ve dnech 4. a 5. listopadu 2010 proběhl ve Vídni v sídle Austrian Institute of Technology „Place, Cultural Heritage and the InternetEuropeanaConnect - Cross-Project Networking Workshop“. Tohoto workshopu (<http://dme.ait.ac.at/workshop2010/>) se zúčastnili nejen zástupci projektu Europeana Connect ale i dalších projektů, které se nějakým způsobem zabývají geografickými informacemi v kontextu kulturního dědictví.

Podrobná zpráva z workshopu je zveřejněna na adrese <http://duha.mzk.cz/blog/zprava-z-workshopu-place-cultural-heritage-and-internet>

Dne 8.11. pak v MZK proběhlo setkání, kde se definitivně uzavřely požadavky na design (viz <http://www.kacerr.cz/mzk/>) nové verze připravovaného webu <http://mapy.mzk.cz/> , který bude plně zpřístupňovat dokumenty naskenované v MZK a také bude demonstrovat geografické hledání a aplikace vytvořené v projektu OldMapsOnline.org. Další jednání pak proběhlo ohledně detailů budoucího vývoje služby Georeferencer. V neposlední řadě také proběhlo osobní setkání a konzultace se studentem FI MU, který v rámci své bakalářské práce pracuje na webové aplikaci využívající WebGL standardu pro zobrazování a transformaci rastrových map přímo ve webovém prohlížeči. Tato bakalářská práce by mohl přispět v budoucnu k lepší vizualizaci map a také zrychlení služby Georeferencer při použití s moderními webovými prohlížeči.

F.2 Grafický návrh pro mapy.mzk.cz



F.3 Statistiky návštěvnosti

Jsou přiloženy jako samostatná příloha.