

Projekt Europeana Newspapers – online brána k evropským historickým novinovým sbírkám



Následující článek, jehož hlavním cílem je krátce představit projekt Europeana Newspapers čtenářům časopisu ITlib a odborné veřejnosti na Slovensku, vznikl hned z několika důvodů. V první řadě je to projekt velice zajímavý a přínosný, což obecně u mezinárodních projektů financovaných z fondů EU nebývá (bohužel) zrovna pravidlem, takže si slovenská veřejnost, vzhledem k faktu, že v projektu nefiguruje žádný slovenský partner, jistě zaslouží být o projektu informována v rámci „československého“ kontextu – čímž se dostáváme k postavě autora článku (který je sice národností české, ale narodil se jako a vždy zůstane Čechoslovák), jemuž se dostalo cti řídit diseminaci a PR projektu Europeana Newspapers v prvním roce jeho realizace, a tak s radostí přijal nabídku připravit o projektu krátký informační článek pro tento časopis.

Veškeré další informace a aktuality projektu mohou případní zájemci nalézt na jeho oficiálních webových stránkách¹, kde se mohou mj. přihlásit k odběru pravidelného Newsletteru.

Pozadí vzniku a hlavní cíle projektu

Nedílnou součástí každého národního kulturního dědictví jsou noviny – zdánlivě zcela obyčejná věc každodenní spotřeby, která nás obvykle přestane zajímat hned druhý den po přečtení, kdy si do nich v lepším případě zabalíme svačinu do práce. Z knihovnického hlediska ovšem staré noviny představují unikátní zdroj historických informací, které vytvářejí velmi komplexní a bohatý odraz dějin, který je vzhledem k obecné povaze médií nutné vždy brát s rezervou a interpretovat v rámci historických souvislostí, což je činnost jako stvořená pro různé historiky, badatele, komentátory či hledače dějinných souvislostí. Před příchodem digitalizace byly historické novinové sbírky vesměs uloženy v temných depozitářích, kde na ně sedal prach, který občas smetl náhodný návštěvník. Dnes se tedy díky novým technologiím otevírá brána k tomuto znovuobjevenému zdroji historických informací, která umožní si pročitat a zkoumat historické noviny komukoliv z pohodlí domova. Na cestě k volné dostupnosti většiny evropských historických novinových sbírek online jsme však teprve na samotném počátku, a jednou z globálních evropských aktivit k naplnění tohoto cíle je i představovaný projekt.

Zasvěcenější čtenáři by mohli jistě namítnout, že většina národních institucí zodpovědných za uchování a poskytování novinových sbírek již v digitalizaci novinových fondů značně pokročilo, či ji některé z nich dokonce úspěšně dokončily, a měli by samozřejmě pravdu. Projekt Europeana Newspapers začíná právě tam, kde končí proces digitalizace v mateřské instituci. Digitalizace totiž neznamená automatické zpřístupnění získaných dat, které se po ukončení procesu digitalizace mohou velice snadno objevit ve stejně zaprášených depozitářích, jako tomu bylo u jejich papírových „matek“, což se v současné době běžně děje. Důvodem je především poměrně finančně i organizačně náročný proces funkčního zveřejnění digitalizovaného novinového obsahu, k jehož vylepšení a zjednodušení by představovaný projekt rád přispěl.

Projekt je tvořen a realizován v rámci partnerství důležitých evropských hráčů v oboru zpřístupňování digitálního novinového obsahu, přičemž na tomto místě zmíníme především Královskou knihovnu v Nizozemí, což je dnes jediná země Evropy,

která zrealizovala online přístup ke svým novinovým sbírkám². V rámci projektu bude předávat své cenné know-how takovým způsobem, aby i ostatní státy mohly v dohledné době zpřístupnit svůj novinový obsah, a to pomocí portálu Europeana³. Nyní se jistě ctěný čtenář ptá, proč je k tomuto zapotřebí mezinárodního projektu, když se každý stát může poučit z nizozemského řešení a naimplementovat nad svým obsahem vlastní řešení? Jediným důvodem jsou – peníze. Nizozemský systém je skvělý, ale jeho implementace byla tak nákladná (i proto, že byla v EU první), že by si jí téměř žádný jiný evropský stát nemohl dovolit. Hlavní cíle projektu v tomto smyslu jsou:

- zpřístupnit v rámci systému Europeana až 18 milionů digitalizovaných novinových stránek s využitím fulltextového vyhledávání,
- vylepšit, zefektivnit a zlevnit procesy vedoucí k transformaci digitalizovaných novinových sbírek v plně využitelný a zpřístupněný digitální obsah,
- standardizace novinových metadat s důrazem na digitalizované zdroje.

Partnerství projektu tvoří především poskytovatelé digitalizovaného obsahu, tedy vesměs národní, resp. státní knihovny (Berlín, Hamburg, Nizozemí, Estonsko, Rakousko, Francie, Finsko, Lotyšsko, Polsko, Turecko, Velká Británie a Jižní Tyrolsko), jako techničtí partneři pak významné univerzitní knihovny (Manchester, Innsbruck, Bělehrad) společně s předním světovým komerčním dodavatelem řešení v oblasti „dolování obsahu“ z digitalizovaných dokumentů (CCS GmbH). Celkem je v projektu 18 partnerů ze 12 evropských zemí včetně Turecka. Česko ani Slovensko v tomto projektu bohužel zastoupeno není, což snad nijak negativně neovlivní budoucí zpřístupnění českých, resp. slovenských historických novinových sbírek – koneckonců, výstupy projektu jsou a budou veřejné (nicméně je škoda prostředků, které v případě partnerství mohly být použity k transformaci několika milionů českých, resp. slovenských digitalizovaných novinových stránek).

Aktivity projektu

I když popisovat strukturu aktivit evropských projektů bývá často poněkud nezábavné, v případě představovaného projektu se tentokrát budeme držet struktury řízení realizace projektu, jelikož je logická a celkem přehledná, takže pomůže čtenářům poměrně složitý proces lépe pochopit. Aktivity projektu jsou rozvržené do 36 měsíců – projekt začal v únoru 2012 – v rámci šesti pracovních balíčků (WP). Celý projekt je pak koordinován Státní berlínskou knihovnou, která je známá svým zřejmě celosvětově nejrozsáhlejším rejstříkem historických novinových titulů (rejstříkem, nikoliv katalogem – ono i komplexně zmapovat, jaké všechny novinové tituly v historii vycházely, je téměř sislyfovský úkol).

WP1 Koordinace a řízení projektu (Státní berlínská knihovna)

Zde jen ve stručnosti poznamenejme, že řídit a zdárně zrealizovat takto náročný mezinárodní projekt není žádná legrace, a bylo nutné aplikovat mnoho metodických i funkčních postupů, včetně

¹ <http://www.europeana-newspapers.eu/>

² Projekt „Historische Kranten“ – dostupný na <http://kranten.kb.nl/>

³ Více viz <http://www.europeana.eu/portal/>

např. online kolaborativního prostředí či důsledného risk managementu. Hlavní cíl balíčku je jasný – úspěšná implementace projektu.

WP2 Rafinace digitalizovaných novin (Královská knihovna Nizozemí)

Rafinace je sice termín vypůjčený z chemie, ale nejlépe se i v češtině hodí na proces „získání a vyčištění kvalitního digitálního obsahu z digitalizovaných materiálů“, který tvoří páteř celé aktivity. Cesta od digitálních souborů naskenovaných novin k portálu, jaký vybudovali v Nizozemí (viz výše), je totiž dlouhá a trnitá. Důvody jsou přitom různé – nedostatečná kvalita skenů, specifická noviny obsah (netradiční fonty, skladba článků a jejich nepravidelná segmentace), finančně i časově náročný proces metody OCR apod. V tomto smyslu má aktivita tyto základní, navzájem posloupné cíle:

- analyzovat digitalizované novinové sbírky partnerů a následně z nich vybrat 18 milionů těch nejlepších pro další rafinaci,
- definovat minimální požadavky na kvalitu vstupního digitalizovaného materiálu pro rafinaci a použití v rámci pokročilých služeb systému Europeana,
- koordinovat a zrealizovat rafinaci vybraných sbírek, s využitím zdokonalených procesů OCR, OLR (Optical Layout Recognition) a NER (Name Entity Recognition),
- poskytnout doporučení o nejlepší praxi procesu rafinace digitalizovaných novinových sbírek.

WP3 – Evaluce a hodnocení kvality (University of Salford)

V podobných komplexních projektech bývá zvykem věnovat celou sadu aktivit neustálému průběžnému vyhodnocování kvality procesů projektu, ať už organizačních, nebo technických. Nejinak je tomu i v projektu *Europeana Newspapers*, kde je vzhledem k jeho povaze dokonce nezbytné navrženou metodiku projektu v jeho průběhu neustále pozměňovat a zdokonalovat. K tomu slouží celá řada unikátních nástrojů vyvinutých na Univerzitě Salford v Manchesteru, která tuto projektovou aktivitu koordinuje. Jejím cílem je především skutečná využitelnost výsledků a výstupů projektu, včetně např. vzhledu a funkcionality budoucí brány (online portálu), pomocí které budou uživatelé v budoucnu přistupovat k digitálnímu novinovému obsahu. Autor článku netuší, jaká je situace na Slovensku, ale v ČR je právě zanedbávání či vyslovená ignorace podobných evaluačních aktivit častou příčinou někdy až do očí bijící nesmyslnosti a nepoužitelnosti výstupů mnoha projektů financovaných z fondů EU.

WP4 – Agregace a zpřístupnění digitalizovaných novin (Královská knihovna Nizozemí)

Poté, co dojde k úspěšné transformaci vybraných digitalizovaných novinových exemplářů do formy digitálních souborů, bude třeba zajistit jejich agregaci pomocí služby *Evropská knihovna*⁴ a jejich následné zpřístupnění v rámci systému Europeana, kde bude pro tyto účely vytvořen speciální novinový portál. Aktivitu znovu vede Královská knihovna Nizozemí, protože jak jsme se již zmiňovali, jako jediná instituce v Evropě má s realizací uceleného národního systému zpřístupnění digitálních historických novinových sbírek zkušenosti. Součástí této aktivity je i zmapování současného stavu v oblasti napříč Evropou (rozsah, povaha, stav digitalizace, stav rafinace, stav zpřístupnění) a vytvoření celoevropského registru digitalizovaných novinových sbírek. Výsledkem pak nebude nic jiného než funkční portál, ve kterém bude možné zdarma fulltextově prohledávat 18 milionů novinových stránek dodaných partnerskými knihovnami projektu.

WP5 – Doporučení nejlepší praxe v oblasti novinových metadat (Univerzita Innsbruck)

Další povahou technická sada aktivit projektu se týká poměrně klíčového aspektu celého procesu rafinace a agregace novinového obsahu, a to jsou metadata. V současné době neexistuje žádný centrální standard pro identifikační popis digitalizovaného novinového obsahu, což značně problematizuje sdílení obsahu či jeho přenos mezi jednotlivými platformami. V tomto směru mají aktivity pracovního balíčku následující dílčí cíle:

- zmapovat a analyzovat existující a v praxi používané metadatové formáty pro digitální novinové kolekce,
- navrhnout a publikovat nový ideální metadatový standard, založený na modelech typu METS, MOTS, ALTO apod.,
- vytvořit online veřejnou znalostní bázi standardu pro libovolné využití knihovnami po celém světě a její úspěšné rozšíření.

WP6 Diseminace a využití výsledků projektu (LIBER)

Posledním pracovním balíčkem projektu je koordinace jeho diseminačních aktivit se zaměřením na odbornou veřejnost, s cílem maximálního budoucího využití výsledků projektu napříč celou Evropou. V tomto smyslu je důležitá systematická propagace projektu v rámci odborné veřejnosti, a to nejen všemi partnery projektu ve svých vlastních geografických oblastech, ale i v nepartnerských zemích (např. pomocí takového článku pomocí tzv. přidružených členů projektu, což jsou vesměs instituce z dalších evropských zemí, které mají v gesci či drží významnou část objemu národního kulturního dědictví v novinové oblasti. Samozřejmostí je aktivní účast projektu na všech významných mezinárodních a vybraných národních konferencích, a to nejen v EU, ale i v zámoří (v roce 2012 bylo i projektu úspěšně přednášeno v Torontu či Washingtonu, což u podobných evropských projektů není zdaleka samozřejmé). V roce 2014 by měla být uspořádána série celkem 10 národních informačních dnů, tedy celodenních akcí v 10 evropských zemích, které budou zaměřeny na propagaci výsledků projektu, a to v národním jazyce. Pro zástupce odborné veřejnosti, především pak držitele digitalizovaných novinových sbírek, jsou pak určeny 3 tematické, chronologicky navazující mezinárodní diseminační workshopy (o rafinaci, o agregaci, o přínosu projektu k agendě Europe 2020).

Shrnutí a závěr

V prostoru vyhrazeném pro představení projektu jsme měli možnost obsáhnout jen zlomek všech zajímavostí a výzev, které projekt přinesl a aktuálně přináší všem partnerským institucím. Kromě zpřístupnění obsahu vzniklého v rámci projektu mohou jeho výsledky přispět ke snadnějšímu a levnějšímu zpřístupnění i dalších sbírek historických novin, a to nejen v zemích EU, ale po celém světě. V současnosti existující nástroje a postupy jsou často nedostatečné, ať už z hlediska jejich kapacity, kvality či nákladů – závěrem můžeme prozradit, že Nizozemí svého cíle dosáhlo nakonec tak, že odeslalo miliony skenů novin do Indie, kde je na zakázku ručně přepsali živí lidé. Pevně věřím, že postupy a výsledky vytvořené v rámci projektu v budoucnu přispějí k realizaci poněkud méně tradičního způsobu transformace historických novinových sbírek v otevřený zdroj informací a inspirace komukoliv s přístupem k internetu.

Aleš Pekárek

pekarek.ales@gmail.com

⁴ <http://www.theeuropeanlibrary.org/tel4/>