



DIGITALIZACE MORAVSKÝCH KNIHOVNÍCH SBÍREK

Moravská zemská knihovna v Brně

**Závěrečná zpráva
leden 2012**

**Zpracovali:
Ing. Petr Žabička
Mgr. Pavla Švástová**

Leden 2012, Brno

Obsah

A Konstatační část.....	3
A1 Rešerše.....	3
A2 Současný stav	3
A3 Vstupní data	6
B Analytická část	8
B1 Vlastní řešení	8
B2 Přínos řešitele	14
B3 Posun znalostí	14
C Návrhová část.....	15
C1 Výsledky řešení	15
C2 Porovnání dosažených výsledků s cíli projektu	16
C3 Uplatněné výsledky v RIV	19
C3.1 Výsledek druhu G - funkční vzorek	19
C3.2 Výsledky druhu B - odborná kniha	19
C3.3 Výsledek druhu D - článek ve sborníku	19
C3.4 Výsledek druhu J - článek v odborném periodiku	20
D Využití účelové podpory	21
D1 Využití účelové podpory v roce 2008	21
D1.1 Účelové běžné výdaje	21
D1.2 Cestovní náhrady	21
D1.3 Mzdové náklady	21
D2 Využití účelové podpory v roce 2009	21
D2.1 Účelové běžné výdaje	22
D2.2 Cestovní náhrady	22
D2.3 Mzdové náklady	22
D2.4 Zákonné sociální pojištění	22
D2.5 Zákonné sociální náhrady	22
D3 Využití účelové podpory v roce 2010	22
D3.1 Výdaje na služby	22
D3.2 Cestovní náhrady	23
D3.3 Mzdové náklady	23
D3.4 Zákonné sociální pojištění	23
D3.5 Zákonné sociální náhrady	23
D4 Využití účelové podpory v roce 2011	23
D4.1 Účelové běžné výdaje	23
D4.2 Cestovní náhrady	24
D4.3 Mzdové náklady	24
D4.4 Zákonné sociální pojištění	24
D4.5 Zákonné sociální náhrady	24
E Stručné resumé, klíčová slova.....	25
E1 Resumé	25
E2 Klíčová slova	25

A Konstatální část

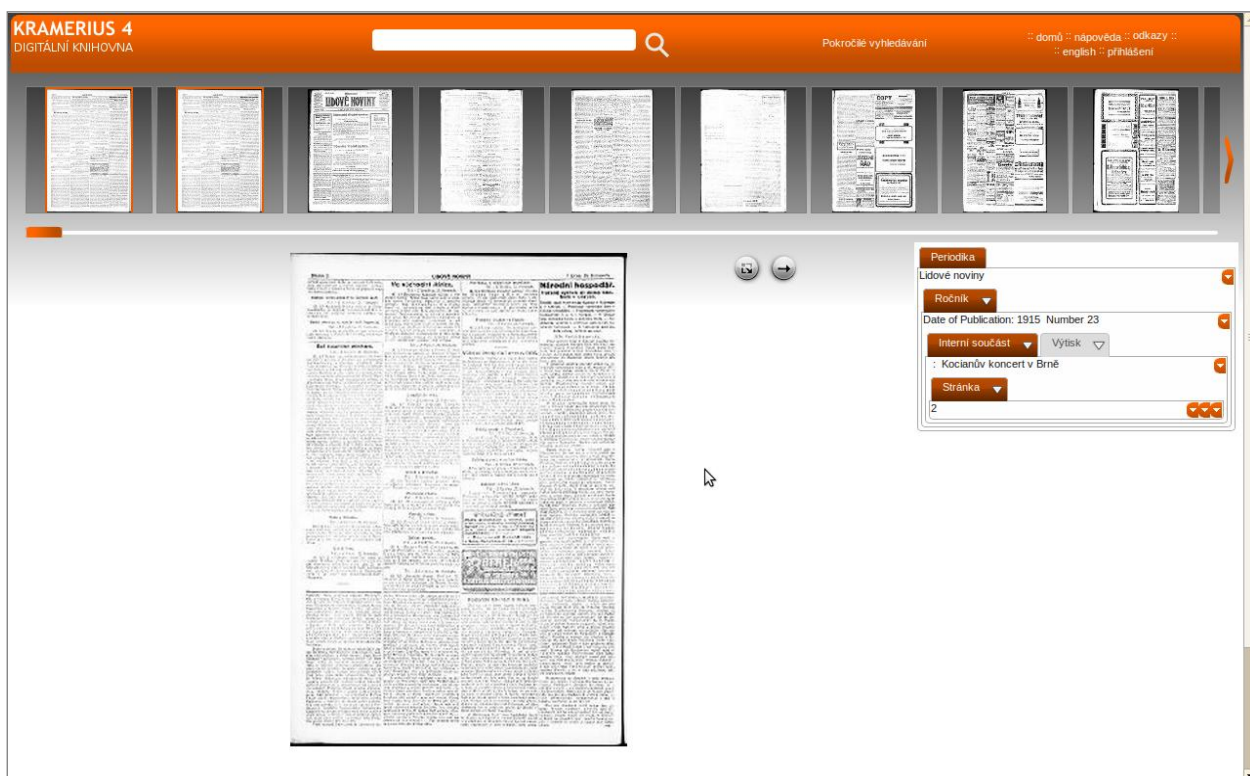
A1 Rešerše

1. Národní digitální knihovna [online]. 13.12.2011 [cit. 2012-01-25]. Dostupné z: <http://www.ndk.cz>
2. [ANL+](#) [online]. 3.10.2011 [cit. 2012-01-25]. Dostupné z: <http://info.jib.cz/news/anl>
3. Matěnová, Růžena. Báze ANL+ a odborné rešeršní požadavky knihovníků a veřejnosti z oblasti hudební vědy. *Ikaros* [online]. 2011, roč. 15, č. 12 [cit. 25.01.2012]. Dostupný na World Wide Web: <<http://www.ikaros.cz/node/7280>>. URN-NBN:cz-ik7280. ISSN 1212-5075.
4. Košťálová, Karolína. ANL+ [online]. 3.-4.11.2011 [cit. 2012-01-25]. Komise pro služby. Dostupné z: www.nkp.cz/files/sluz_komise2011Kostalova.pps
5. Kramerius [online]. 26.1.2012 [cit. 2012-01-29]. Dostupné z: <http://code.google.com/p/kramerius/>

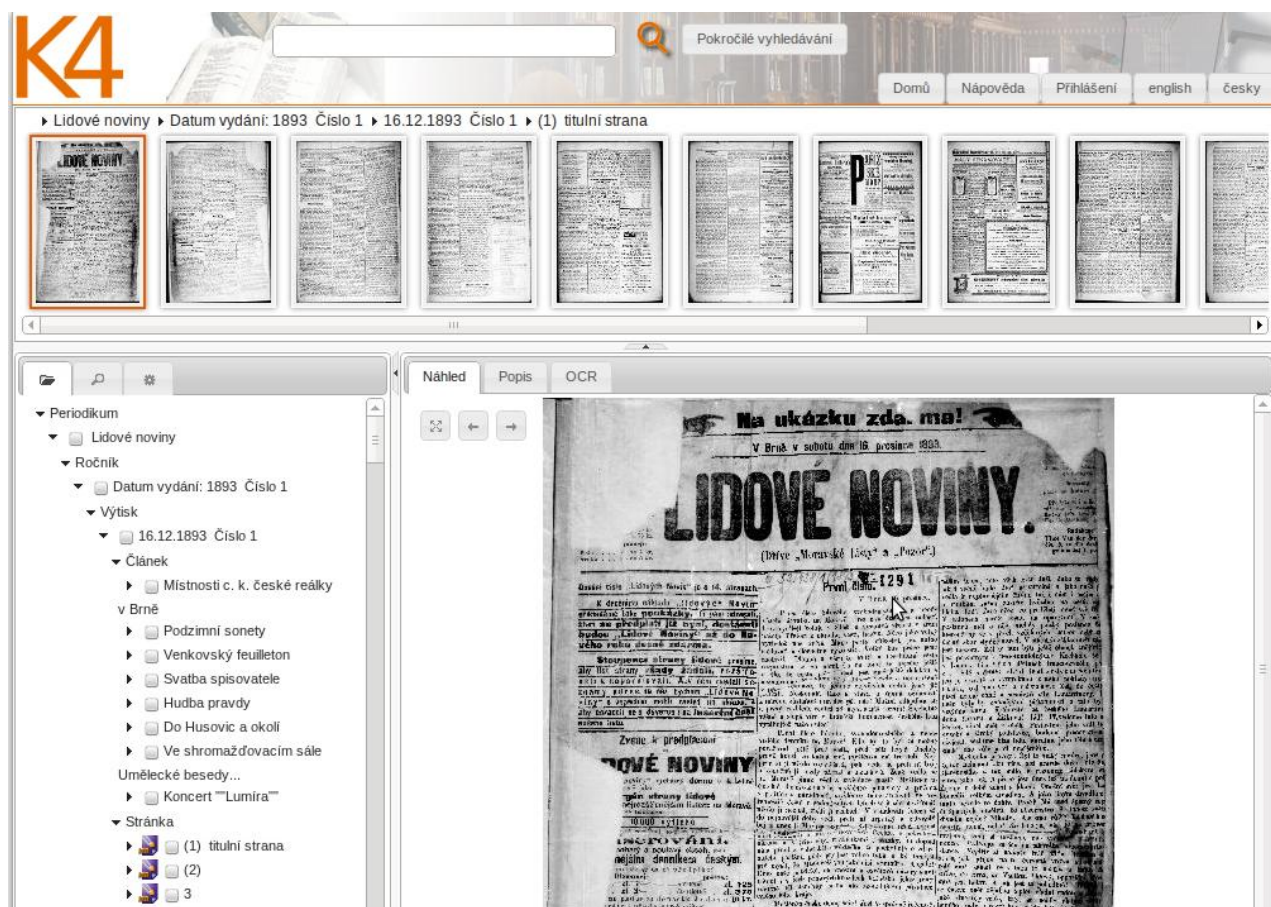
A2 Současný stav

Projekt Digitalizace moravských knihovních sbírek má návaznost na řadu dalších projektů Moravské zemské knihovny (dále MZK) i dalších knihoven v ČR. Spolu s Národní knihovnou (dále NK) spolupracujeme na projektu Národní digitální knihovna (NDK), který byl schválen, proběhlo výběrové řízení na systémového integrátora, který by měl zajistit dodání veškeré techniky, hardwaru a softwaru, jejich dodání na místo určení, zapojení, nainstalování a vzájemné integrace. Začátek provozu se v MZK předpokládá na polovinu roku 2012. Výsledky projektu Digitalizace moravských knihovních sbírek by měly být uloženy k archivaci v LTP systému, který bychom měli získat v rámci projektu NDK.

Pro zobrazování článků využíváme poslední stabilní verzi digitální knihovny Kramerius 4, která je založena na digitálním repozitáři Fedora Commons. Aktuální ostrá verze je v provozu na adrese <http://kramerius.mzk.cz/>. Digitální knihovna Kramerius prodělala v posledním roce a půl výrazné změny. Od první betaverze, která sice již podporovala zobrazení článků, ale nebylo možné v nich efektivně listovat a vyhledávat, třídít apod. až k současné verzi, která je již pro zobrazení a vyhledávání článků lépe uzpůsobena. Proběhla integrace imageserveru, který umožňuje zobrazovat obrázky JPEG 2000 externě, bez nutnosti mít je naimportované přímo v repozitáři Fedora Commons. Výhodou je optimalizace času importu, uživatelsky přívětivější prohlížení velkoformátových dokumentů bez nutnosti stahovat plug-in, jak je nyní nutné pro obrázky ve formátu DJVU (např. Lidové noviny plánujeme zkonvertovat z DJVU do JPEG2000). Vývoj v této oblasti s koncem tohoto projektu samozřejmě nekončí. Ačkoli systém primárně vyvíjí Knihovna akademie věd spolu s Národní knihovnou ČR, MZK s těmito knihovnami spolupracuje a systém intenzivně provozně testuje a přispívá i k rozvoji jeho funkcionality. S dalšími verzemi se budeme soustředit další zlepšení uživatelského rozhraní i funkcionality.



Obrázek č. 1: Kramerius verze 4.3



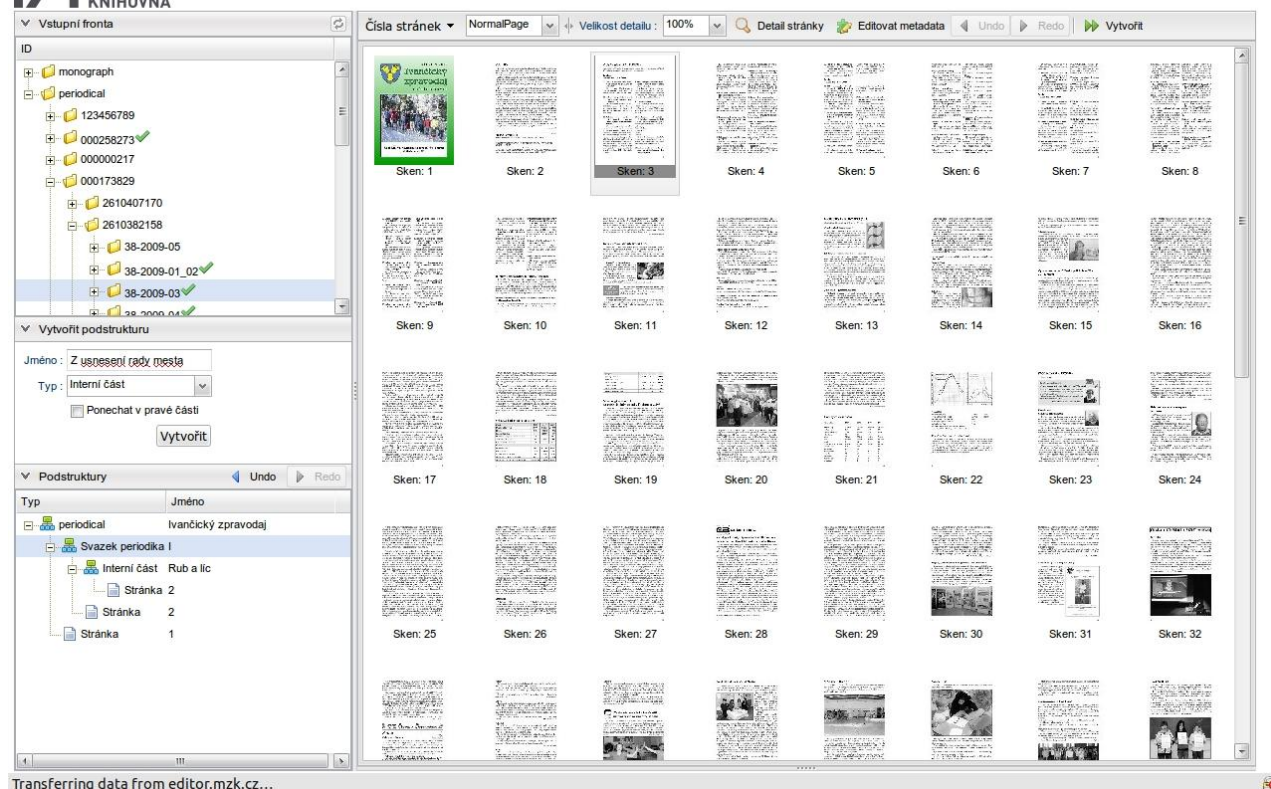
Obrázek č. 2: Kramerius verze 4.4.1 nasazený v Moravské zemské knihovně

S prezentací zdigitalizovaných dat dále souvisí projekt Registr digitalizace (www.registrdigitalizace.cz), do kterého je MZK aktivně zapojena od roku 2009. Registr digitalizace je společným projektem Národní knihovny ČR a Knihovny Akademie věd, jeho realizaci zajišťuje firma Incad. Základní myšlenkou projektu bylo shromáždit informace o dokumentech digitalizovaných na našem území na jediné místo, aby se zabránilo zbytečným duplicitám při stále narůstajících objemech digitalizovaných dat a umožnilo tedy i synchronizaci masové digitalizace v MZK a NK. V první řadě je určen pro knihovny (a další instituce), které digitalizují nebo to mají v plánu. Zjistí zde, zda dokument, který se chystají zpracovat, nedigitalizoval již někdo jiný, s vlastníkem dokumentu se mohou domluvit, zda by neposkytl kopii pro zveřejnění ve vlastní digitální knihovně apod. Knihovny, které mají vybudovanou digitalizační linku, mohou využít Registr digitalizace i pro evidenci vlastních digitalizačních aktivit a jako nástroj pro řízení procesu digitalizace (digitalizační workflow). Další vývoj navazuje na projekty Národní digitální knihovna a Kramerius verze 4, proběhl přechod na novou vnitřní datovou strukturu, která umožní lepší navázání na metadata z knihovních katalogů a digitální knihovny Kramerius tak, aby bylo možné většinu procesů řešit automatizovaně. Zároveň by mělo být lépe řešeno přidělování perzistentních identifikátorů novým digitálním objektům, což je důležité např. právě pro články.

Vytvářet článková metadata a publikovat je v Krameriu verze 3 lze pomocí metadatového editoru vyvinutého v MZK v projektu hrazeném z VISKu 3 Obálky knih a metadatový editor. V současnosti se tento editor v MZK rutinně využívá při interní digitalizaci. Z důvodu přechodu na Krameria verze 4 byl zahájen vývoj nové verze metadatového editoru, který bude podporovat metadatové formáty MODS a FOXML a umožní editaci a vytváření metadat právě pro tuto verzi. Je ve fázi testování a v průběhu února 2012 by měl přejít do provozního testování a nahradit tak původní editor.

The screenshot shows the 'Metadata Editor' interface. At the top, there's a navigation bar with links like 'search', 'statistics', 'authors', 'verify', 'help', and 'admin'. Below this, the title 'Metadata Editor' is centered. The main area is titled '--- Undefined article name ---'. It contains a form with various fields for metadata entry, organized into sections like 'Unique Identifier', 'Core Bibliographic Description Periodical', 'Title', 'Title Group', 'Sorting Title', 'Key Title', 'Codens', 'GMD', 'Publisher', 'Printer', 'Physical Description', 'Series', 'Language', 'Subject', and 'Keyword'. Each field has a dropdown menu for selection and a small 'x' icon for clearing the field. The 'Title' section is expanded, showing 'Main Title' with the text 'Propowjedy a podobienstwj'. To the right of the form is a preview of the document, showing a page from a manuscript with the title 'Propowjedy a podobienstwj.' and several paragraphs of text in a historical script. The preview includes a 'prev | next | references' navigation bar and a 'OCR [1/1]' indicator.

Obrázek č. 3: Metadatový editor pro Krameria 3 – popis článků



Obrázek č. 4: Metadatový editor pro Krameria 4 – vytváření článků

Národní knihovna řeší od loňského roku projekt ANL+. Jedná se o experimentální zdroj poskytující informace o článcích publikovaných v českých novinách, časopisech a dalších pokračujících zdrojích v roce 2011. ANL+ navazuje na databázi ANL, kterou společně vytvářely od roku 1992 Národní knihovna České republiky, Moravská zemská knihovna v Brně, krajské knihovny a odborné knihovny. Záznamy článků a náhledy jsou v ANL+ volně dostupné, přístup k plným textům a digitalizované podobě článků je možný pouze z vybraných počítačů zúčastněných knihoven. Základem této skupiny knihoven jsou knihovny, které mají v rámci projektu VISK8/A zajištěn přístup k databázi firmy Anopress. Do konce roku 2011 byla báze ANL+ dostupná v rozhraní Primo, kam jsme pro testování posílali též naše článkové metadata z Lidových novin. Úspěchem bylo, že data bylo možné bez problémů zobrazit a prolinkovat na plné texty.

A3 Vstupní data

V rámci příprav na projekt byl vybrán soubor kulturně významných historických moravských periodik a monografií, který měl být zdigitalizován a článkově zpracován.

Zpracovávané dokumenty:

1. *Mährisches Magazin* (Brno, 1789) ve třech svazcích vyjadřoval moravské vlastenectví, byl kritický k pobělohorské době, nové vyd. z r. 1792 svědčí o zajímavosti obsahu
2. *Allgemeines europäisches Journal* (Brno, 1794-1798) měsíčník, vyhledávány často pro ilustrace Mozartových děl
3. *Povídatel* (1815 – 1817) seriál z doby národního obrození od Josefa Matěje Sychry
4. *Moravsko-slezský časopis pro lid* (Brno, 1842-1846) první český měsíčník na Moravě

určený pro venkovský lid, red. F. Diebl. Titul byl již zpracován v rámci přípravy projektu.

5. *Týdenník* (Brno, 1848) první moderní český týdeník na Moravě, red. J. Ohéřal
6. *Nový život* (Nový Jičín-Prostějov, 1896-1907) umělecký měsíčník Katolické moderny, red. K. Dostál-Lutinov
7. *Večery* (Brno, 1911-1914) týdeník vydávaný jako příl. Lidových novin přinášel původní beletristickou tvorbu, red. A. Heinrich
8. *Moravsko-slezský sborník* (Ostrava, 1918-1920) politická a kulturní revue o vzniku ČSR, red. V. Martínek
9. *Pásmo* (Brno-Praha, 1924-1926) mezinárodní moderní měsíčník levicového Devětsilu, red. F. Halas, B. Václavěk
10. Jako doplněk časopisů bude rozepsaná kniha Vincence Brandla *Kniha pro každého Moravana* (Brno, 1863), která je první českou vlastivědnou publikací země. Analyticky byla zpracována topografická část, aby bylo možné získat plný text k jednotlivým obcím.
11. Posledním a nejrozsáhlejším zpracovávaným titulem byly Lidové noviny (Brno, 1893-1952), které byly excerpovány výběrově, přesto však poskytly nejrozsáhlejší vstupní data pro další práci.

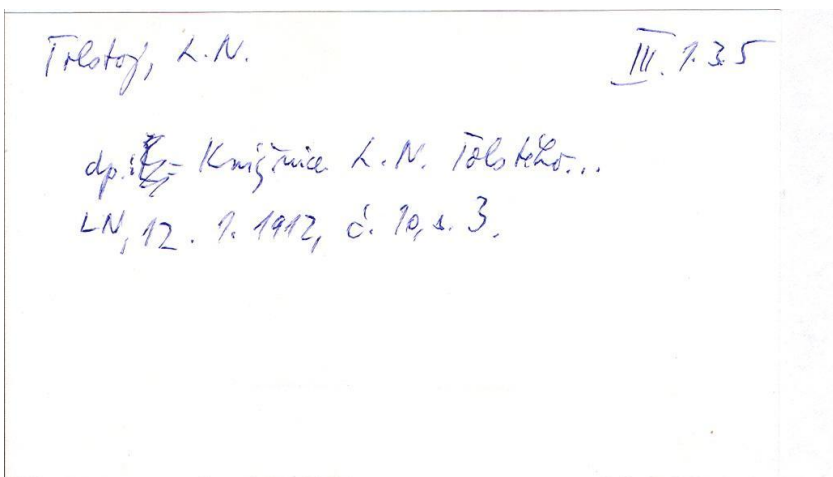
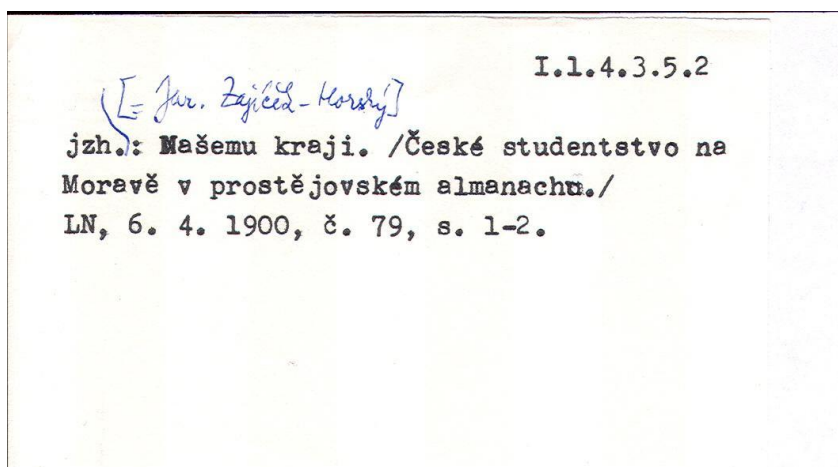
Data vzniklá excerpací prvních devíti titulů byla publikována tiskem a elektronicky, u posledních dvou titulů pak pouze elektronicky.

B Analytická část

B1 Vlastní řešení

V roce 2008 probíhala zejména excerpce nových záznamů z Lidových novin a přepis nebo zpracování záznamů, které již byly zpracovány dříve na knihovních lístcích. Při převodu záznamů se postupovalo dvěma způsoby:

1. Na ručně psané záznamy nebylo možné využít metody OCR, proto byly přepsány do požadované struktury do tabulky .xls.
2. Psacím strojem pořizené bibliografické záznamy byly naskenovány a převedeny do textu nástrojem OCR, ručně upraveny a zformátovány.



Obrázky č. 5 a 6: Ukázky naskenovaných katalogizačních lístků

Na převod do textu OCR byl použit ABBYY FineReader. Vzhledem k nízké kvalitě textu a také k tomu, že na záznamech byly často ručně psané poznámky k obsahu záznamu, nebyl výsledek převodu vždy dokonalý a bylo nutno každý záznam po převedení do textu upravit podle naskenovaného originálu a teprve poté strukturovat do jednotlivých položek.

Vinou špatné kvality papíru i psacího stroje se často špatně převáděly znaky s diakritikou, převodník generoval nové řádky i tam, kde na záznamu nebyly apod. Z toho důvodu a pro množství ručně psaných poznámek na záznamech nebylo možno převedené texty strukturovat automatizovaně, ale proces vyžadoval ruční úpravu a doplňování záznamů před následným tagováním. Mezi jednotlivé oblasti popisu byly vloženy znaky tabulátoru, dále byl záznam překopírován do připraveného formátu .xls tabulky a tabulátory v textu zajistily, že údaje se uložily do správných sloupců. V obou případech zpracování, při ručním přepisu i převodem OCR do textu byl výstupem strukturovaný formát záznamu v připravené .xls tabulce, která obsahuje sloupce: věcné třídění, autor, název článku, podnázev, koho se článek týká, bibliografická citace. Postup byl časově náročný a vyžadoval pozornost pracovníka, ale vzhledem ke kvalitě katalogizačních lístků byl přijat jako jediný možný. Záznamy uložené do tabulky byly ještě zkontrolovány, zejména v oblasti bibliografických citací.

Jedním z cílů **na rok 2009** bylo vytvořit program, který metadata článků z excelových tabulek převede do XML ve formátu, který lze vložit do systému Kramerius verze 3. Z analýzy vyplynulo několik problémů, které souvisí s dosavadním systémem vytváření záznamů v tabulkách.

č.	A	B	C	D	E	F	G	H	I	J
č.	žánr	autor	název	podnázev	popis	roč.	datum	čís.	vyd.	str.
1	1	Slaměník, František	Ludvík Kellr mrtev		Předseda berlínské Společnosti Komenského - nekrolog	23	19150322	74	ranní	9
2	3	Digamma [=František Votruba]	Ze Slovenska		Fejeton	23	19150310	69	ranní	2
3	3	Dr. Š.	Jak se uzdravuji		Fejeton	23	19150326	78	ranní	1
4	3	št. [=Bohuslav Štěchovský]	V novinách nad čarou...		Fejeton	23	19150403	86	odpol.	2
5	3	-is [=Josef Hais?]	Polární zář		Fejeton	23	19150302	61	ranní	2-3
6	5		Poslední Mohykáni		Reakce na fejeton Obec pensistů z 21. 4.	23	19150505	117	ranní	1-2
7	3	Kompiš, Peter ml.	Jako učeň dedil synov		Povídka	25	19170114	12	ranní	1-2
8	3	Rameau, Jean	Mařenčin román		Román na pokračování	23	19150409- 19150702	91- 175	odpol.	5-6, 3
9	3		Alois Kulíšek, domobranec			23	19150328, 19150404, 19150418, 19150425, 19150504, 19150516	80, 87, 93, 100, 107, 116, 128	ranní, ranní, ranní, ranní, ranní, ranní	1-2, 1-2, 1-2, 1-2, 1-2, 1-2

Obrázek č. 7: Ukázka tabulky

Každý řádek tabulky reprezentuje jeden článek a obsahuje metadata deskriptivní a strukturální. V deskriptivních metadatech není jednotně řešeno jméno autora ve slouci B – někdy je ve tvaru „Příjmení, Jméno“ (např. Příkryl, Ondřej), někdy je uveden pseudonym, pod kterým psal do novin a jméno ve tvaru [=„Jméno Příjmení“] (např. -is [=Josef Hais?]), proto není možné jednoduše rozčlenit do jednotlivých polí v XML. Navíc dosavadní krameriovské XML neumožňuje zadat alternativní jméno autora, takže přicházíme v mnoha případech o pseudonym či jinou formu jména autora.

Další problém je ve sloupci E, který stručně popisuje, o co se ve článku jedná a slohový útvar. Z žánrů jsou nejčastěji zastoupeny recenze, fejetony, referáty, sloupky a úvahy. Data nejsou zcela jednotná, takže jsme se zatím rozhodli vložit celý text jako obsah prvku *anotace*. Podrobnější rozčlenění na žánry je možné, ale nepřesné, např. v řádku 6 v tabulce výše se nachází text „reakce na fejeton“, pokud by probíhalo rozčlenění na žánry podle pravidla: „pokud najdeš v textu slovo „fejeton“, přidej článku atribut „fejeton“, neodpovídalo by to v tomto

případě pravdě, jelikož se nejedná o fejeton, ale pouze reakci na něj. Navíc formát pro periodika pro Krameria 3 nepočítá se všemi uvedenými žánry.

Bohužel nebyl na počátku projektu systémově vyřešen zápis strukturálních metadat v případech, kdy se jedná o článek na pokračování. Reálně existují dvě možnosti:

1. Článek vycházel na pokračování v souvislé řadě – potom je datum ve slouci G ve tvaru „od_data-do_data“ a číslo ve sloupci H ve tvaru „od_číslo-do_číslo“ (viz tabulka, řádek 8)
2. Článek není v souvislé řadě – pak jsou uvedena všechna data oddělená čárkami ve sloupci G a k nim příslušný počet čísel atd. (viz tabulka, řádek 9)

Z těchto dat obecně nebylo složité vyextrahovat strukturální metadata a na základě nich přiřadit deskriptivní metadata k příslušnému ročníku, číslu a stránce, která jsou uložena v digitální knihovně. V tabulce se však vyskytují nepřesnosti a chyby, kvůli kterým nelze článek v plném textu spolehlivě dohledat. Např. v řádku 8 je to sice souvislá řada čísel, ale uvedeny jsou dvě reprezentace strany, není poznat, které číslo strany patří ke kterému číslu periodika. V řádku 9 zase nesouhlasí počet dat a čísel.

Na základě analýzy bylo rozhodnuto, že se využije vývoje Krameria verze 4, který bude na základě našich poznatků počítat s importem článkových metadat a jejich napojením na plné texty. Bibliografická metadata budou ve formátu MODS a struktura bude určena nativním formátem úložiště Fedora Commons – FOXML, provazba mezi stránkami je zajištěna RDF vazbami. Byl zadán požadavek na rozvoj systému Kramerius 4 – propojení Metadatového editoru připraveného v rámci projektu MZK a pilotní rozšíření struktury v systému Kramerius pro ukládání článků – popis článků v rámci stávající MODS. Ve spolupráci s odborníky z Národní knihovny byla vytvořena analýza popisu pro články ve formátu MODS, která obsahuje doporučení na základě zahraničních specifikací s ohledem na data excerpcovaná z Lidových novin i popisu článků v ANL.

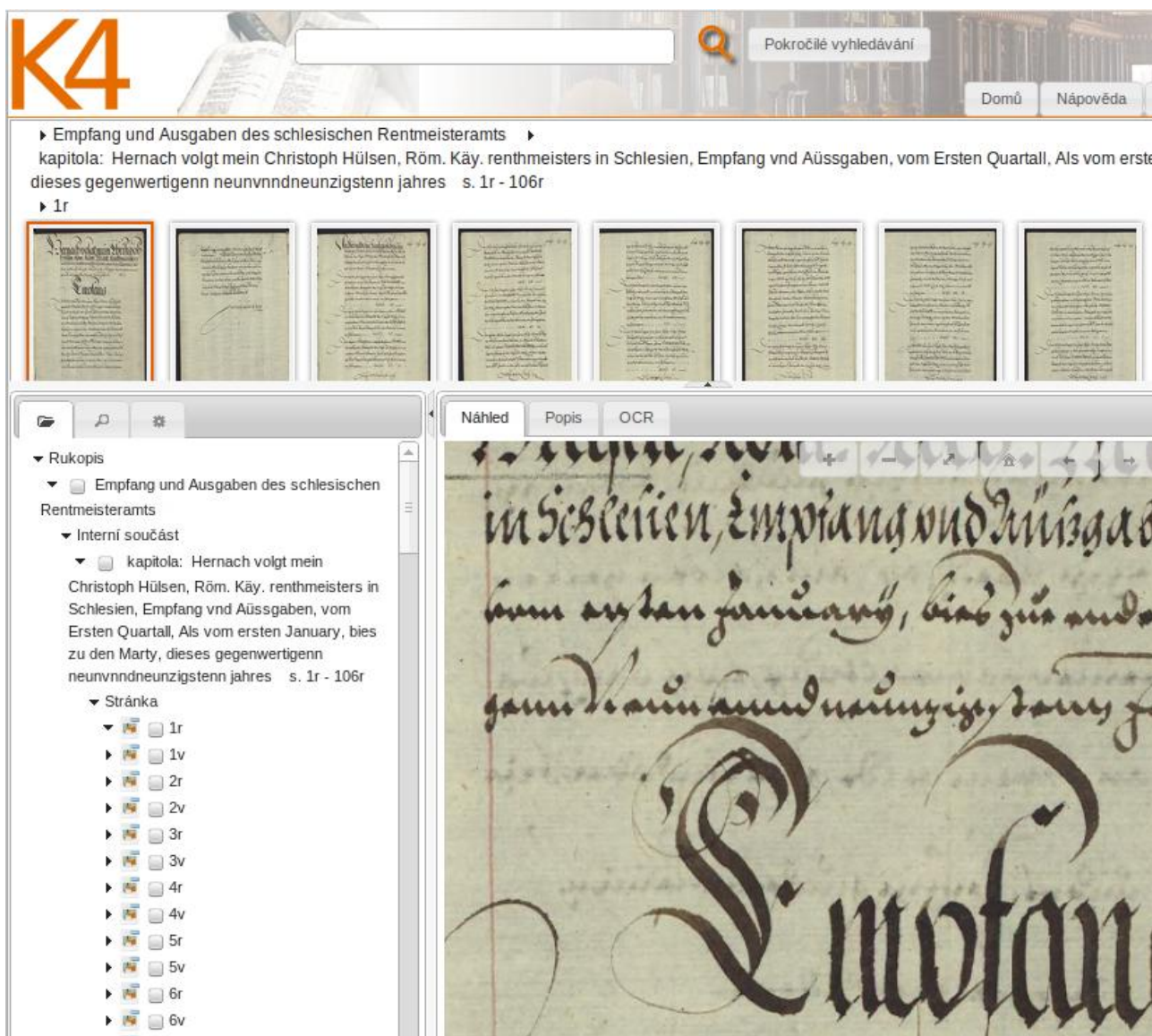
V roce 2010 byla provedena zásadní změna v řešitelském týmu. Hlavním řešitelem se stal místo Doc. Dr. Jaromíra Kubíčka, CSc. Ing. Petr Žabička. Hlavními důvody byl přesun důrazu z bibliografických činností na technologickou část projektu a odchod původního řešitele do důchodu.

Byly ukončeny práce na excerpcích. Celkově bylo vytvořeno 63 193 záznamů článků, největší objem z Lidových novin.

titul	počet záznamů
Lidové noviny (z let 1901-1926, 1939-1945)	57 381
Mährisches Magazin	24
Allgemeines europäisches Journal	926
Moravsko-slezský časopis pro lid	623
Týdenník	484
Večery	828
Moravsko-slezský sborník	283
Pásmo	298
Kniha pro každého Moravana	603
Nový život	1 743
Moravsko-slezská revue	1 092
Index	1 137

Byl vytvořen program, který metadata článků z excelových tabulek převádí do XML ve formátu, který lze vložit do systému Kramerius verze 4. Program je vyvíjen jako open-source pod licenci GNU-GPL v3 a je umístěn na <http://code.google.com/p/kramenius4-data-imports/> (kromě programu na kontrolu, konverzi a import článků obsahuje nástroje pro zpracování, konverzi a import rukopisů a dalších dokumentů). Hromadnou analýzou strukturálních metadat byla zjištěna nekonzistence dat v některých řádcích tabulky. Typickým problémem bylo, že nesouhlasil počet stran a počet příslušných čísel u článků na pokračování, chybějící označení, zda se jedná o ranní či odpolední vydání, nejednotné oddělovače a nechtěné překlapy. Chybovost je zapříčiněna ve většině případů lidským faktorem a dále ne zcela vhodným návrhem původní struktury excelové tabulky, která díky malé granularitě popisu a nutnosti používat jednotné oddělovače takovéto chyby dovoluje. Z analyzovaných dat byly zpětně vytvářeny nové tabulky, které obsahují chybné nebo nejednoznačné záznamy z původních tabulek. Byl vytvořen validační nástroj, který prověřil tabulková data a ta bezproblémová vyřadil k dalšímu zpracování - konverzi a importu do Krameria. V dalším kroku byla hledána nějaká opakující se chyba, která se ve zbytku nezpracovatelných záznamů objevovala a byla strojově hromadně opravitelná. Tato část dat byla opravena a vyloučena z chybových záznamů a následovalo hledání další systematické chyby atd. Tato procedura zabrala několik měsíců prací včetně konverzí dat a testování. Zbylo několik tisíc záznamů, které obsahují chyby, které již není možné opravit jinak než ručně s originálním dokumentem v ruce. Obecně lze však říct, že vzhledem k celému objemu všech zpracovaných článkových metadat je chybovost poměrně nízká.

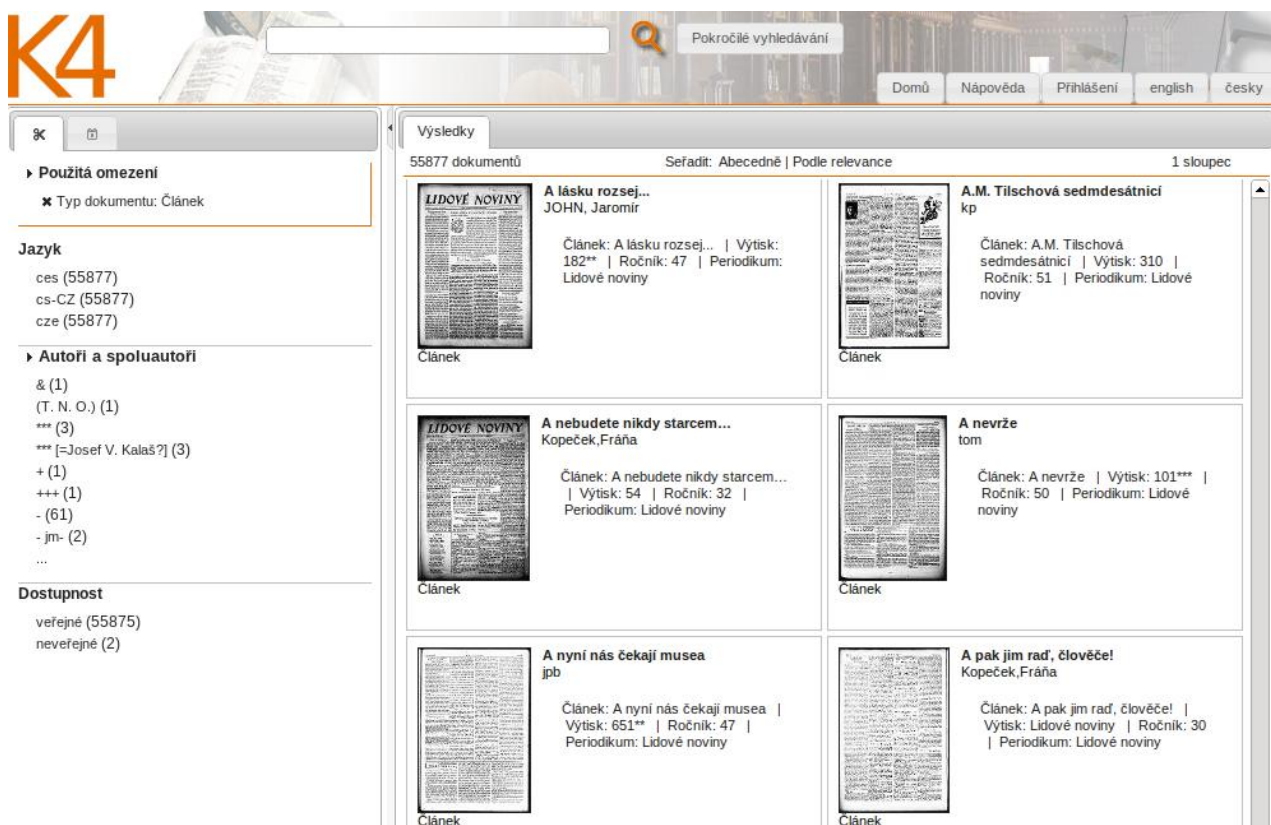
První vzorek konvertovaných metadat byl importován do demoverze Krameria 4, kde byly články označeny jako „interní součást“. Chybovost a nestandardnost tabulek a hledání způsobu, jak z nich vytěžit co nejvíce záznamů způsobila mírné zdržení v importu kompletních dat do Krameria 4. Kromě toho jsme v této době zápasili s různými problémy Krameria 4 - nefunkční indexace, chyby při importu apod. Po překonání těchto obtíží byla většina dat naimportována do Krameria 4 Moravské zemské knihovny (<http://kramenius.mzk.cz>), kde jsme pro ně vytvořili samostatný modul „Články“ respektive „Articles“.



Obr. č. 8: Ukázka dokumentu zobrazeného přes imageserver – možnost zoom přímo v okně

Byla upravena funkčnost metadatového editoru pro Krameria 3 pro výrobu článkových metadat a byl otestován import a zobrazení. Vzhledem k tomu, že obrázky a článková metadata nevznikala současně, upustilo se od výroby článků vyráběných v rámci tohoto projektu přímo v editoru.

V roce 2011 probíhaly závěrečné práce na importu článkových metadat do Krameria 4 a jejich zobrazení. Dalo by se říct, že to byl rok ve znamení testování, optimalizace zobrazení a prací na odstraňování chyb. Zobrazení v Krameriu verze 4.3 nám nevyhovovalo z mnoha důvodů. Nelze v něm smysluplně vyhledávat a prohlížení je v počtu našich cca 60 tisíc záznamů de facto nemožné, protože bylo možné pouze posouvat se v abecedně seřazeném seznamu jen o deset článků vpřed či vzad. V Krameriu verze 4.4, která má nový vzhled a vylepšené administrátorské funkce, jsou již články prohlédatelné, lze v nich vyhledávat a listovat lépe, lze je třídit podle více kritérií, omezit časové období na časové ose apod.



Obr. č. 9: Zobrazení článků v Krameriu 4.4.1

Použitelnost samotných metadat byla ověřena v testovacím projektu ANL+, který zobrazoval excerpované články v prostředí Primo. Balík importních metadat ve formátu FOXML s vnořeným popisem v MODS byl obohacen o odkaz do Krameria a zaslán k testování. Přestože import a praktické testy proběhly úspěšně, pro oficiální prezentaci byla data Lidových novin prozatím ze zpřístupnění prostřednictvím tohoto projektu vyjmuta, protože tvořila se zbytkem dokumentů časově nesourodý celek a byla tak pro uživatele matoucí.

Metadatový editor pro Krameria 4 je ve fázi testování a lze v něm, stejně jako v jeho předchůdci, vyrábět článková metadata a publikovat je.

Soubor článkových metadat se nachází kromě Krameria 4 v SVN systému MZK. SVN je verzovací systém, podobně jako CVS nebo GIT. Soubory v SVN serveru se nacházejí v několika repozitářích a udržují se předchozí verze (=revize) těchto souborů. Typicky se používá na zdrojový kód, ale v MZK se tímto systémem verzují XML soubory. Výhodou je právě skutečnost, že systém drží předchozí verze souborů – pokud jsme tedy upravovali několikrát celý balík článkových metadat nebo pouze části či jednotlivé záznamy, máme všechny jejich verze a známe historii změn. Právě tento soubor metadat spolu s konverzním programem vytvořeným v rámci projektu a reálným vystavením záznamů v digitální knihovně považujeme za funkční vzorek, který tak není pouze statickým souborem záznamů, ale je z něho možné zjistit vývoj záznamu a je přístupný online.

B2 Přínos řešitele

V prvních letech projektu byl kladen důraz hlavně na vybudování datového základu v rozsahu vyexcerpovaných bibliografických záznamů, v převodu klasicky psaných záznamů na katalogizačních lístcích do strojem čitelné podoby konverzí OCR a v tištěných publikačních výstupech.

Po excerpování dostatečného množství záznamů se přistoupilo k jejich analýze a zjišťování toho, jak se která metadatová schémata s takovým popisem vyrovnají. Upustilo se od plánu importovat metadata do Krameria verze 3 a pozornost byla přesunuta k vývoji Krameria verze 4, který v průběhu řešení projektu začal vznikat. Vzhledem k pravidelným účastem na schůzkách věnovaných jeho vývoji jsme měli možnost využít naše dosavadní znalosti a zkušenosti z analýz a zahrnout je do hlavní větve vývoje. Pro Kramerijs verze 4 byla vytvořena specifikace formátu MODS (nejen) pro článková metadata, která se stala součástí národní verze specifikace pro periodika vydané na konci roku 2011. Mimo to jsme přispěli k debatám při vzniku projektu ANL+ a pro testování jsme poskytli veškeré naše záznamy.

Dalším přínosem projektu byla analýza a následná implementace image serveru IIPImage do Krameria verze 4. Image server řeší doposud problematické a poněkud uživatelsky nepřívětivé zobrazování větších formátů dokumentů, jako jsou například mapy nebo právě stránky novin. V budoucnu to může otevírat cestu k využití anotačních nástrojů vyvíjených v projektu Europeana Connect, které jsou založeny na technologii OpenLayers.

S článkovými metadaty dlouhodobě počítáme i ve vývoji metadatového editoru - dříve pro Krameria 3, nyní nového editoru pro Kramerijs 4. V obou editorech lze článková metadata vytvářet, editovat i publikovat do příslušné verze Krameria.

B3 Posun znalostí

Řešitelé v prvním roce trvání projektu si ověřovali časovou náročnost stanoveného projektu na prvním vzorku vytipovaných titulů určených k analytickému zpracování. Výsledky ukazovaly na reálnost projektu a to jak u významných časopisů z 18.–20. století, tak u deníku *Lidové noviny*. Práce na excerpování tedy pokračovaly a byly ze zpracovaných časopisů vydávány monografické publikace.

V průběhu dalších let projektu byly operativně řešeny problémy, které vyplynuly z analýzy vytvářených článkových metadat. Ideální by byla změna v systému vyplňování excelové tabulky – zavést větší granularitu při vyplňování – každý údaj do zvláštního sloupce, logičtější rozdělení strukturálních metadat apod., aby byla tato data lépe strojově zpracovatelná. V době, kdy začaly analytické práce, však bylo vyexcerpováno takové množství záznamů, že bylo výhodnější mít balík jednotně zpracovaných dat.

Pro budoucí zpracování by bylo nejlepším řešením zapisování příslušných metadat přes metadatový editor vyvíjený MZK, díky tomu budou články okamžitě navázány na příslušný sken stránky jejich publikování v Kramerijs 4. Nespornou výhodou zapisování článků přímo přes metadatový editor je eliminace chyb ve strukturálních metadatech. Takové chyby se v excelové tabulce vyskytují a jsou určující pro to, zda se příslušný metadatový záznam sváže s konkrétním skenem stránky, na kterém se příslušný článek nachází. Neovlivní to však překlepy a chyby v popisných metadatech, které vedou k tomu, že článek není v takovém případě vyhledatelný.

C Návrhová část

C1 Výsledky řešení

V textu projektu je cíl projektu specifikován takto: "Cílem řešení projektu je digitalizace vybraného souboru seriálů, zpracování jejich formálních a obsahových metadat se záměrem zefektivnit a zkvalitnit přístup k bohatému a mezinárodně vyhledávaným informačním zdrojem pro oblasti literární historie, kultury a poznání země. Projekt bude řešen na osmi titulech časopisů, 1 titulu novin (články o kultuře a krásná tvorba v deníku Lidové noviny) a 1 knize. Kromě analytického popisu půjde o zpracování metodiky tvorby a zpřístupnění obsahových metadat a jejich vyhledávání na regionálním datovém úložišti."

V rámci projektu byly zpracovány tituly:

Lidové noviny (Brno, 1893–1952) deník, analytickým rozpisem byly zpracovány kulturní rubriky a původní beletristická tvorba. Vzniklo cca 60 tisíc záznamů článků, z nich 55 877 se podařilo automatizovaně zpracovat a dle aktuálních doporučení vytvořit metadata a vložit do Krameria 4. Kompletní Lidové noviny zdigitalizované a analyticky zpracované se nachází zde: <http://kramerius.mzk.cz/search/handle/uuid:56775c82-435f-11dd-b505-00145e5790ea>

Allgemeines europäisches Journal (Brno, 1794-1798) měsíčník, vyhledávaný často pro ilustrace Mozartových děl. Analytické zpracování bylo vydáno jako monografie: <http://kramerius.mzk.cz/search/handle/uuid:8069c22a-0ab2-41c3-b7a5-5dcca16ce088>

Povídatel (1815–1817) seriál z doby národního obrození od Josefa Matěje Sychry. Byl zdigitalizován, vystaven je zde: <http://kramerius.mzk.cz/search/handle/uuid:adb1e6c9-d852-11e0-876b-0050569d679d>

Moravsko-slezský časopis pro lid (Brno, 1842-1846) první český měsíčník na Moravě určený pro venkovský lid, red. F. Diebl. Titul byl již zpracován v rámci přípravy projektu. Analytické zpracování bylo vydáno jako monografie a časopis byl zdigitalizován: <http://kramerius.mzk.cz/search/handle/uuid:1f77cb0c-9e0e-11e0-a742-0050569d679d>
<http://kramerius.mzk.cz/search/handle/uuid:fb0dd004-91b1-11e0-bf0c-0050569d679d>

Týdenník (Brno, 1848) první moderní český týdeník na Moravě, red. J. Ohéral. Časopis byl zdigitalizován a je vystaven zde: <http://kramerius.mzk.cz/search/handle/uuid:efb06369-efdd-11e0-96d0-0050569d679d>

Nový život (Nový Jičín-Prostějov, 1896-1907) umělecký měsíčník Katolické moderny, red. K. Dostál-Lutinov. Seriál byl zdigitalizován i analyticky zpracován: <http://kramerius.mzk.cz/search/handle/uuid:b75722a2-935c-11e0-bdd7-0050569d679d>
<http://kramerius.mzk.cz/search/handle/uuid:927e5ac7-9e19-11e0-a742-0050569d679d>

Večery (Brno, 1911-1914) týdeník vydávaný jako příloha Lidových novin přinášel původní beletristickou tvorbu, red. A. Heinrich. Seriál byl zdigitalizován i analyticky zpracován: <http://kramerius.mzk.cz/search/handle/uuid:5e98c820-451e-11de-b499-000d606f5dc6>
<http://kramerius.mzk.cz/search/handle/uuid:7b04cf9b-9e80-11e0-b8ae-0050569d679d>

Moravsko-slezský sborník (Ostrava, 1918-1920) politická a kulturní revue o vzniku ČSR, red. V. Martínek. Časopis byl zdigitalizován: <http://kramerius.mzk.cz/search/handle/uuid:40e2f4dd-91ad-11e0-bf0c-0050569d679d>

Pásmo (Brno-Praha, 1924-1926) mezinárodní moderní měsíčník levicového Devětsilu, red. F. Halas, B. Václavek. Časopis byl analyticky zpracován: <http://kramerius.mzk.cz/search/handle/uuid:e864bc9e-9e21-11e0-a742-0050569d679d>

Jako doplněk časopisů byla rozepsána kniha Vincence Brandla ***Kniha pro každého Moravana*** (Brno, 1863), která je první českou vlastivědnou publikací země. Tato publikace byla článkově rozepsána ve formě indexu, který je vystaven na stránkách projektu. Položky v indexu vedou na konkrétní stránky digitalizovaného dokumentu v Krameriu 4. Index se nachází zde:

<http://www.mzk.cz/o-knihovne/vyzkum-projekty/narodni-projekty/digitalizace-moravskych-knihovnich-sbirek/kniha-pro>

Digitalizovaná monografie se nachází zde:

<http://kramerius.mzk.cz/search/handle/uuid:88be4507-b393-11e0-b43e-0050569d679d>

C2 Porovnání dosažených výsledků s cíli projektu

- ⤴ *“Staré digitalizované dokumenty, zvláště pak noviny a časopisy, stojí v tomto směru stranou zájmu. Náš projekt si klade za cíl začít tuto pomyslnou mezeru vyplňovat a jeho realizace přispěje k lepšímu využití vybraného souboru publikací významných z hlediska kulturně-historického a geopolitického.”*

Tento cíl byl zcela naplněn. Byly zdigitalizovány a excerpovány významné časopisy a noviny z 19. století. Viz předchozí kapitola.

- ⤴ *“Cílem je “analytické zpracování obsahu uvedeného vybraného souboru publikací, resp. novin a časopisů. Předpokládaný rozsah je soubor 60 tisíc záznamů, který představuje dost rozsáhlý komplex jednotlivých typů dokumentů jak pro svůj cenný obsah, tak pro další metodiku zpracování.”*

V rámci projektu bylo vyexcerpováno více než 60 tisíc záznamů z výše zmíněných periodik. Samostatná metodika nevznikla, ale je součástí specifikace pro periodika. Tato metodika je zveřejněna zde: <http://ndk.cz/upravy-ve-specifikaci-metadat-pro-periodika-verze-1.2>

- ⤴ *“Spolupráce bibliografů se systémovými knihovníky a programátory má pak za cíl zpracovat snadno přístupný a uživatelsky příznivý systém, který přes obsahová data umožní vyhledávat plné texty. Půjde o práci se systémem Kramerius pro vyhledávání obsahu, příp. dalšími systémy v této oblasti.”*

V Krameriu verze 4.4.1 je naimportováno téměř 60 tisíc článků Lidových novin, lze v nich vyhledávat, je možné je procházet, třídit, omezovat výběr fazetovým vyhledáváním či pomocí časové osy.

- ⤴ *“Předpokládá se i ověření a možnosti spolupráce s profesionální firmou, která dosud digitalizaci periodik pro MZK externě provádí, metadata však zachycuje jen formální, ne obsahová. Projekt ověří stávající technologie digitalizace včetně užívaných metadat v zahraničí.”*

V současnosti vznikají ve firmě Elsys pro projekt ANL+ metadata pro periodika včetně formátu ALTO, který obsahuje OCR a zónování jednotlivých článků na stránce. Podle toho by bylo možné navázat článek nejen na příslušný ročník, číslo a stranu či rozsah stran, ale přímo na zónu na stránce, kde se článek nachází. Toto je zatím v počátcích, nicméně v MZK bylo prakticky ověřeno, že Kramerius 4 v současné verzi již import formátu ALTO podporuje a dokáže s ním i pracovat při zobrazení vyhledávaných termínů na naskenované stránce, i když zatím jen pokud je tato stránka ve formátu jpg.

- ⤴ *“Počítá se s pomocí a zapojením odborných pracovníků Národní knihovny ČR, kteří jsou tvůrci koncepce digitalizace pro ČR, a s propojením na odborné týmy, řešící tuto problematiku v zahraničí.”*

Spolupracovali jsme s pracovníky Národní knihovny na národní specifikaci pro periodika a

monografie jako výstup z digitalizace projektu Národní digitální knihovna a VISK7. Podílíme se na vývoji Krameria verze 4.

- ▲ *“Systém Kramerius bude díky probíhajícímu vývoji v průběhu tohoto roku připraven na práci s částmi textu - články a kapitolami. Díky tomu bude možné již od počátku možné tato metadata importovat do Krameria a využívat je pro vyhledávání. Aby však byla zajištěna možnost spolehlivě se na tyto dokumenty odkazovat, bude nutné do Krameria doplnit podporu pro mezinárodně používané perzistentní identifikátory (např. DOI), které umožní badatelům i široké veřejnosti odkazovat na jednotlivé články s jistotou, že daný odkaz bude i po delší době funkční. Pravděpodobně bude také nutné vylepšit podporu Krameria pro aktualizace údajů v jeho interní databázi tak, aby tento proces bylo možné snadno provázat se systémem pro pořizování záznamů článků v reálném čase.”*

Původní hypotéza, že se budou článková metadata importovat do Krameria verze 3, se nepotvrdila. Po úvodních analýzách a testech jsme se rozhodli počkat, zapojit se do vývoje Krameria 4 a prosadit možnost importu a zpřístupnění zde. Nemusely se tak “ohýbat” procesy a funkce Krameria 3. V Krameriu 4 má každý článek svůj identifikátor UUID, přes který je možné na něj jednoznačně odkazovat.

- ▲ *“Přínosem projektu bude jednak obohacení digitalizovaných sbírek o soubor významných moravik, jejich analytický rozpis, ale zejména zpracování metod pro provádění obsahových metadat podle standardů, schopných snadného a přehledného vyhledávání.”*

Výše v kapitole C1 je uveden seznam periodik včetně odkazů do digitální knihovny Kramerius, kde jsou veřejně přístupné.

- ▲ *„Přínos v aktuální problematice digitalizace knihovnických fondů v paměťových institucích“*

MZK se významnou částí podílí na vývoji nových SW nástrojů, které jsou využívány v oblasti digitalizace dokumentů. Jednak je to projekt Kramerius verze 4, kde byla v roce 2009 iniciována integrace článkových metadat a v roce 2010 byl implementován image server IIPImage. Dále je testováno zobrazování obrázků v nestandardních velikostech a formátech. Doposud se v digitální knihovně využíval jako prezentační formát pro user copy DJVU a JPEG, v rámci projektu Národní digitální knihovna začnou vznikat JPEG2000 soubory, které dokáže IIPImage poměrně rychle a uživatelsky přívětivě zobrazit.

V projektu Registr digitalizace jsme se podíleli na přechodu na novou vnitřní datovou strukturu, která byla předtím vystavěna hlavně pro potřeby Národní knihovny a Knihovny akademie věd. Snažili jsme se o to, aby byla více univerzální a počítala i s jinými, poněkud nestandardními typy dokumentů – mapami, grafikami, starými tisky a rukopisy, hudebninami apod. V současnosti se řeší hlavně automatizace procesů sklizení metadat z digitálních knihoven apod. Pravidelně poskytujeme firmě Incad, která Registr digitalizace vyvíjí, testovací data a na testování se podílíme hlavně lidskými zdroji.

- ▲ *“Posouzení stávajících SW pro zpracování obsahových metadat u knih a článků z časopisů a novin, jejich dopracování a optimalizace po ověření k vypracování postupů k provozní praxi”*

V rámci České republiky nebyl k dispozici SW pro výrobu článkových metadat kromě Metadatového editoru pro Krameria 3 vyvíjeného v Moravské zemské knihovně. V tomto SW bylo možné vytvářet a publikovat články do Krameria 3, ale pouze k nově zpracovávaným časopisům. Vzhledem k tomu, že na počátku projektu bylo rozhodnuto, že digitalizace a zpracování článků bude probíhat odděleně (Metadatový editor v té době nebyl plně funkční a zavedený do digitalizační praxe), nebylo možné zpracovávat články průběžně z digitálních kopií. Lidové noviny byly zdigitalizovány již dříve z projektu VISK7 a z ostatních časopisů nebyla článková metadata

zapisována do excelové tabulky, ale byly z nich vydány monografie, jejich digitalizace probíhala až později.

▲ *“Tištěné publikační rejstříky osmi excerpovaných vytipovaných časopisů z 18.–20. století”*

1. KUBÍČEK, Jaromír. Moravsko-slezský časopis pro lid 1842-1846. Brno : Moravská zemská knihovna, 2007. 41 s. Rejstříky moravského tisku. Sv. 1. ISBN 978-80-7051-169-5 (brož.). Dostupné z: <http://kramerius.mzk.cz/search/handle/uuid:1f77cb0c-9e0e-11e0-a742-0050569d679d>
2. PAPÍRNÍK, Miloš. Magazin nützlicher und angenehmer Lektüre aus verschiedenen Fächern für den denkenden Leser aus allen Ständen 1782 ; Mährisches Magazin 1789. Brno : Moravská zemská knihovna, 2008. 20 s. Rejstříky moravského tisku. Sv. 2. ISBN 978-80-7051-174-9. Dostupné z: <http://kramerius.mzk.cz/search/handle/uuid:5a711f33-9dfb-11e0-a742-0050569d679d>
3. KUBÍČEK, Jaromír. Večery : beletristická příloha Lidových novin 1911-1914. Brno : Moravská zemská knihovna, 2008. 51 s. ISBN 978-80-7051-175-6. Dostupné z: <http://kramerius.mzk.cz/search/handle/uuid:7b04cf9b-9e80-11e0-b8ae-0050569d679d>
4. PAPÍRNÍK, Miloš. Allgemeines europäisches Journal 1794–1798. Brno : Moravská zemská knihovna, 2009. 82 s. Rejstříky moravského tisku. Sv. 4. ISBN 978-80-7051-182-4. Dostupné z: <http://kramerius.mzk.cz/search/handle/uuid:8069c22a-0ab2-41c3-b7a5-5dcca16ce088>
5. KUBÍČEK, Jaromír. Týdeník, listy poučné a zábavné 1848–1849. Brno : Moravská zemská knihovna, 2009. 36 s. Rejstříky moravského tisku. Sv. 5. ISBN 978-80-7051-183-1. Dostupné z: <http://kramerius.mzk.cz/search/handle/uuid:1be27f32-2d10-11e0-b59b-0050569d679d>
6. BUKOVÁ, Eva. Nový život 1896 – 1907. Brno : Moravská zemská knihovna, 2009. Rejstříky moravského tisku. Sv. 6. ISBN 978-80-7051-184-8. Dostupné z: <http://kramerius.mzk.cz/search/handle/uuid:927e5ac7-9e19-11e0-a742-0050569d679d>
7. KUBÍČEK, Jaromír. Pásmo 1924-1926. Index 1929-1939. 1. vyd. V Brně : Moravská zemská knihovna, 2010. 104 s. ISBN 978-80-7051-185-5. Dostupné z: <http://kramerius.mzk.cz/search/handle/uuid:e864bc9e-9e21-11e0-a742-0050569d679d>
8. KUBÍČEK, Jaromír. PAPÍRNÍK, Miloš. Moravsko-slezská revue 1905-1923. Moravsko-slezský sborník 1918-1920. 1. vyd. V Brně : Moravská zemská knihovna, 2010. 167 s. ISBN 978-80-7051-188-6. Dostupné z: <http://kramerius.mzk.cz/search/handle/uuid:1ef76c9c-9e0d-11e0-a742-0050569d679d>

C3 Uplatnění výsledky v RIV

C3.1 Výsledek druhu G - funkční vzorek

Funkční vzorek "Analytická bibliografie moravských periodik" umožňuje zpracování a zpřístupnění velkého množství záznamů článkové bibliografie online. Testuje nové postupy tvorby článkové databáze a integraci této databáze do digitálních knihoven. Vstup dat do systému využívá optimalizovaných tabulek s údaji důležitými pro identifikaci článku. Data jsou transformována do XML a importována do digitální knihovny, kde jsou propojena s plnými texty. Na základě testování byla vyrobena stabilní verze. Funkční vzorek umožňuje zefektivnit proces tvorby metadat, čímž dochází k úspoře osobních nákladů. Je využíván v MZK a je volně k dispozici dalším institucím.

[RIV/00094943: /10:#0000078 - Analytická bibliografie moravských periodik \(2010\)](#)

<http://www.isvav.cz/h11/resultDetail.do?sessionId=E59D97D2431E5410C5AF6537158299CA?rowId=RIV%2F00094943%3A%2F10%3A%230000078!RIV11-MK0-00094943>

C3.2 Výsledky druhu B - odborná kniha

KUBÍČEK, Jaromír. PAPÍRNÍK, Miloš. Moravsko-slezská revue 1905-1923. Moravsko-slezský sborník 1918-1920. 1. vyd. V Brně : Moravská zemská knihovna, 2010. 167 s. ISBN 978-80-7051-188-6.

[RIV/00094943: /10:#0000054 - Moravsko-slezská revue 1905-1923 Brünner Wochenblatt 1824-1827 \(2010\)](#)

<http://www.isvav.cz/resultDetail.do?rowId=RIV%2F00094943%3A%2F10%3A%230000054!RIV11-MK0-00094943>

KUBÍČEK, Jaromír. Pásmo 1924-1926. Index 1929-1939. 1. vyd. V Brně : Moravská zemská knihovna, 2010. 104 s. ISBN 978-80-7051-185-5.

[RIV/00094943: /10:#0000061 - Pásmo 1924-1926; Index 1929-1939 \(2010\)](#)

<http://www.isvav.cz/resultDetail.do?rowId=RIV%2F00094943%3A%2F10%3A%230000061!RIV11-MK0-00094943>

PAPÍRNÍK, Miloš. Allgemeines europäisches Journal 1794–1798. Brno : Moravská zemská knihovna, 2009. 82 s. Rejstříky moravského tisku. Sv. 4. ISBN 978-80-7051-182-4.

[RIV/00094943: /09:#0000060 - Allgemeines europäisches Journal 1794-1798 \(2009\)](#)

<http://www.isvav.cz/resultDetail.do?rowId=RIV%2F00094943%3A%2F09%3A%230000060!RIV11-MK0-00094943>

BUKOVÁ, Eva. Nový život 1896 – 1907. Brno : Moravská zemská knihovna, 2009. Rejstříky moravského tisku. Sv. 6. ISBN 978-80-7051-184-8.

[RIV/00094943: /09:#0000050 - Nový život 1896-1907 \(2009\)](#)

<http://www.isvav.cz/resultDetail.do?rowId=RIV%2F00094943%3A%2F09%3A%230000050!RIV10-MK0-00094943>

C3.3 Výsledek druhu D - článek ve sborníku

KUBÍČEK, Jaromír. K problematice české bibliografie. In: *Sdružení knihoven České republiky. Rok 2009*. Brno: Sdružení knihoven České republiky, 2009, s. 72-75. Dostupné z: http://www.snk.sk/swift_data/source/NBU/Zborniky/BZ_2008_2009/Kolokvium_Zvolen_2008/Kubicek_Jaromir.pdf

Článek popisuje vývoj bibliografických pracovišť v českých zemích od dob, kdy sestavovala bibliografické soupisy každá městská knihovna do současnosti, kdy se jim věnují některé veřejné knihovny a vědecké instituce. Detailněji se autor popisuje souběžnou českou bibliografií, retrospektivní českou bibliografií, personální bibliografií a souběžnou oborovou bibliografií. V současnosti se bibliografický záznam stává součástí popisovaného elektronického dokumentu, plných textů.

[RIV/00094943: /09:#0000043 - K problematice české bibliografie \(2009\)](#)

C3.4 Výsledek druhu J - článek v odborném periodiku

KUBÍČEK, Jaromír. Německý tisk v českých zemích, jeho stav a zpracování. *Vlastivědný věstník moravský*. 2009, **61**(4), 362-369. ISSN 0323-2581.

Článek informuje o historii vydávání německého tisku na území Čech a Moravy od počátků v 17. století až do druhé světové války. Příspěvek chce ale poukázat na možnosti využití bibliografie při digitalizaci periodik, jejíž předností je kompletní nabídka přístupu a ucelená představa možností pro vyhledávání zpracovaných informací. V závěru je uveden přehled bibliografické práce v okolních zemích, tj. Rakousko, Slovensko, Polsko a je představen projekt Moravské zemské knihovny Digitalizace moravských knihovních sbírek.

[RIV/00094943: /09:#0000042 - Německý tisk v českých zemích, jeho stav a zpracování \(2009\)](#)

D Využití účelové podpory

E Stručné resumé, klíčová slova

E1 Resumé

V projektu Digitalizace moravských knihovních sbírek byla řešena problematika získávání článkových metadat z historicky významných periodik a jejich následné automatické zpracování. V první fázi projektu bylo odbornými bibliografy vytvořeno na 60 tisíc záznamů článků z kulturně významných moravských periodik. Excerpované články byly potom v druhé fázi projektu automatizovaně zpracovány a převedeny do standardizovaných formátů pro import do Krameria 4.

Zkoumaly se hlavně možnosti metadatových formátů, zpřístupnění získaných a konvertovaných metadat a efektivní propojení na již existující data v digitální knihovně Kramérius. V současné době ve verzi 4.4.1 bylo vylepšeno uživatelské rozhraní pro prohlížení a vyhledávání v článkových metadatech a byl implementován image server IIPImage, který umožní efektivnější zobrazování stránek novin v lepších rozlišeních a nových formátech.

Výsledky projektu byly podkladem pro národní specifikaci metadatových formátů, resp. část zabývající se článkovými metadaty. Projekt je přínosný i v dalších aktivitách českých i zahraničních knihoven v oblasti digitalizace – projektech Národní digitální knihovna, Kramérius 4 a ANL+, teoretické poznatky byly využity i v projektech souběžných či navazujících, řešících problematiku identifikátorů digitálních objektů či dlouhodobého uchovávání digitálních dokumentů.

E2 Klíčová slova

Digitalizace, digitální knihovny, systém Kramérius verze 4, informační technologie, knihovní sbírky, bibliografické texty, článková bibliografie, metadata, články, Registr digitalizace, Národní digitální knihovna, ANL+