



Centrální portál knihoven

Petr Žabička, Petra Žabičková
Moravská zemská knihovna v Brně

Elektronické služby knihoven III.

Proč vlastně CPK?

- Jedna z hlavních priorit Koncepce rozvoje knihoven
- Cíl: poskytnout nabídku knihoven na jednom místě uživatelsky přívětivým způsobem
- CPK je nástrojem OPRAVDOVÉ integrace
 - SK ČR ani JIB uživatelům nestačí
 - Uživatelé chtějí vyšší kvalitu služeb
- Jsou k dispozici vhodné nástroje a zároveň jsou knihovny nakloněny úzké spolupráci

Aktuální situace

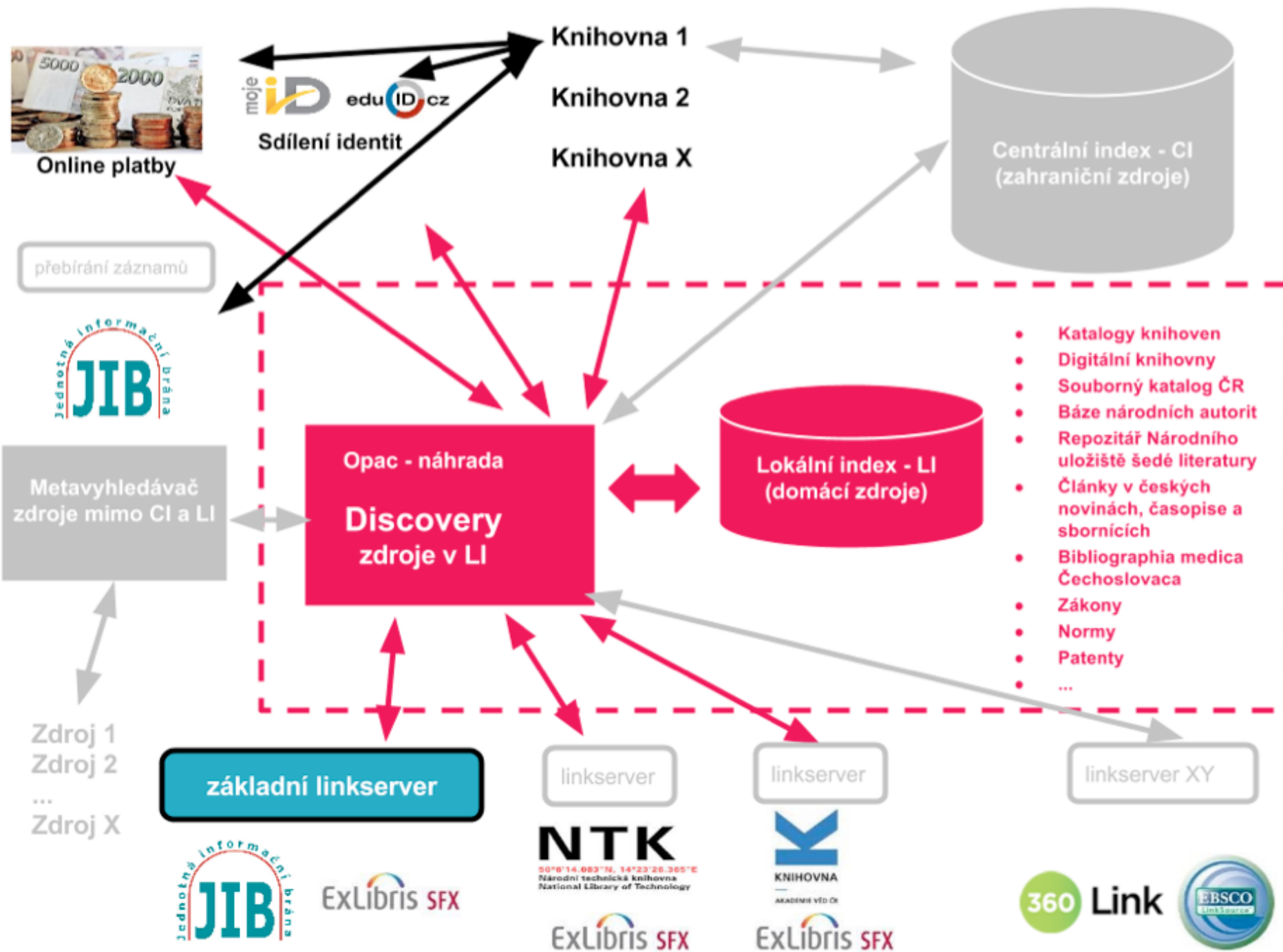
MZK byla na sklonku roku 2014 pověřena Radou CPK a ÚKR vývojem pilotní verze CPK v roce 2015 (financování z VISK8B).

MZK tedy vystupuje ve dvou rolích:

- jako knihovna zapojující se do CPK
- jako dodavatel softwarového řešení.
 - Přitom vychází z podkladů, které připravily Rada pro CPK a pracovní skupiny pro CPK.

Cíle pro rok 2015

- Vybudovat pilotní verzi portálu
- Na bázi systému VuFind 2
- Prioritou zapojení katalogů a dalších zdrojů pilotních knihoven
- Nenahradí JIB ani oborové brány
- Využije stávající linkovací servery (SFX apod.)
- Bude integrovat základní služby knihovnických katalogů tam, kde to bude možné
- Bude obsahovat data ze SK ČR
- Data z dalších zdrojů tam, kde to bude možné



Klíčové komponenty

Aplikace:


- VuFind verze 2.x
- Fulltextová indexace v Apache SOLR
- Správce stažených záznamů
- Data - zdroje lokálního indexu
- Externí linkovací servery
- API zapojených knihovnických systémů

Zakládající knihovny

Rozhodnutí o tom, že se knihovna zapojí, je zcela na každé knihovně a je to rozhodnutí dobrovolné.

Cílem MZK je nalézt technické řešení a zapojit ty knihovny, které chtějí a splňují základní technické požadavky.

Co chceme letos dosáhnout

- sklízení dat z knihovních systémů a digitálních knihoven (ČDK), další vybrané zdroje
- deduplikace a obohacování o dodatečné informace
- zprovoznění běžných funkcí VuFindu
- integrace služeb knihoven / náhrada OPACu
- napojení na linkovací servery jednotlivých zapojovaných knihoven + SFX server JIB
- zapojení jako service provider do  edu ID cz

Aktuální stav implementace

- Příprava od počátku 2015
- Plný rozjezd v březnu
- Připravujeme infrastrukturu
- Dokončujeme novou verzi aplikace pro sklízení, obohacování a deduplikaci záznamů
- Připravujeme prostředí pro zapojení služeb knihoven (NCIP protokol pro knihovny nedisponující API)
 - Ladí se konfigurace pro jednotlivé knihovní systémy

Aktuální stav implementace

Digitální knihovny:

- Předpokládáme agregaci na straně ČDK a sklízení tohoto zdroje

Články

- Předpokládáme sklízení ANL, případně i další zdroje metadat článků

Souborné databáze

- Souborný katalog ČR
- Úvahy o využití informací ze SK VPK

Stahování a správa záznamů

Původní Record Manager byl zcela přepracován

- php -> java - stabilita, rychlost
- stahování primárně přes OAI-PMH, možno dávkově
- všechny stažené záznamy uchovávány beze změn a virtuálně propojovány
- deduplikace a příp. obohacení záznamů probíhá zde, před indexací v SOLRu

Deduplikace záznamů

vychází z algoritmu používaného v portálu finna + modifikace (ve vývoji)

- sestavení deduplikačních klíčů
 - název, autor, rok vydání, typ dokumentu apod.
 - pro podobnost názvu rozhodující Levenshteinova vzdálenost
- na základě klíčů stanovení kandidátů na deduplikaci
- rozhodnutí o tom, které z “podezřelých” záznamů sloučit

Sloučené záznamy

- Pro fulltextovou indexaci vznikne tzv. sloučený záznam (sjednocení všech údajů slučovaných záznamů)
- Pro fasety se použije nejkvalitnější záznam
- Pro zobrazení se používá vždy původní záznam z některé knihovny, nikdy sloučený záznam
- Výběr toho, který záznam se zobrazí, rozhodne algoritmus na základě informací o uživateli a dalších údajích

Náhrada lokálního OPACu - rozsah

- Ověření identity uživatele
- Zobrazení jednotek včetně aktuální dostupnosti jednotky
- Objednání/rezervace publikace
- Profil uživatele / zobrazení
 - jméno, příjmení
 - adresa, emailová adresa
 - platnost registrace
 - blokace (přestupky)
 - příp. skupina/typ uživatele
- Seznam aktuálních výpůjček (a prodloužení)
- Seznam objednávek a rezervací
- Seznam pokut a poplatků

CPK vs. knihovní systémy

- CPK jen zprostředkuje služby dané knihovny - nenahrazuje výpůjční systém, katalogizaci ani akvizici.
- Drží si pouze kopii bibliografických záznamů a údajů o exemplářích.
- Nedrží si informace o uživatelích (výjimky viz. dále), ani o jimi čerpaných službách
 - Ty pouze zprostředkovává ve chvíli, kdy si je uživatel sám aktivně vyžádá.

Technické požadavky

- OAI-PMH set pro všechny veřejné záznamy, včetně expandovaných údajů o jednotkách v poli 996
- Shibboleth jako identity provider / eduID
- NCIP nebo Aleph RESTful API

<http://www.knihovny.cz/o-projektu/jak-se-zapojit-do-projektu/>

OAI

- Úvodní jednorázová sklizeň všech záznamů
 - sklizeň MZK trvá cca 3 hodiny
- 1x denně v noci budeme stahovat všechny změny, které se udály od poslední sklizně
- Máme zájem o všechny změny, které v záznamech provádíte, včetně změn v jednotkách!
- Nevadí nám, když provedete hromadnou změnu a my budeme muset stáhnout většinu záznamů automaticky znovu.

Shibboleth a identity uživatelů

- Každá plně zapojená knihovna se musí zapojit do federace identit eduID (www.eduid.cz)
- Databáze čtenářů knihovny se stane zdrojem identit (identity provider) v rámci eduID
- Technické řešení pro knihovny s NCIP prostřednictvím brány NCIP - Shibboleth vyvinuté v MZK
- CPK bude poskytovatel služby - service provider
- eduID technicky zajišťuje možnost propojit vzájemně identity uživatele z více institucí

Informace o uživateli v portálu

Uživateli, který se do portálu přihlásí, bude při prvním přihlášení vytvořen v portálu účet.

O uživateli držíme ty informace, které si standardně drží VuFind:

- login
- jméno a příjmení
- id lokálního knihovního systému
- datum prvního a posledního přihlášení
- seznamy oblíbených, komentářů, tagů, které si uživatel v portálu sám vytvoří

Všechny ostatní věci přitékají do portálu až ve chvíli aktivní činnosti uživatele

Aleph RESTful API / NCIP

- Rozhraní pro vzájemnou komunikaci dvou systémů
 - CPK jako klient (ptá se)
 - každá knihovna jako server (odpovídá - XML)
- Komunikace přes http/https
 - Aleph: <https://developers.exlibrisgroup.com/aleph/apis/Aleph-RESTful-APIs>
 - Ostatní systémy: NCIP 2.02 <http://www.ncip.info/the-standard.html>

Zabezpečení

- přesný rozsah dotazů, které CPK bude vůči vašemu knihovnímu systému podávat na vyžádání dodáme
- omezit Aleph RESTful API/NCIP server vaší knihovny jen na IP adresu CPK (nutno nastavit na vašem firewallu)
- silně doporučujeme nasazení **https** místo http
 - možnost využít bezplatné certifikáty poskytované CESNETem
- Apache reverzní proxy s overováním jména a hesla ?

Harmonogram

Sklízení katalogů začne brzy, k dalším zdrojům se budeme dostávat postupně

Zapojování služeb bude probíhat postupně

- V tuto chvíli ladíme a připravujeme napojení zástupců jednotlivých knihovnických systémů
- Jakmile odladíme alespoň jednu knihovnu od daného systému, začneme zapojovat další
- Možná dojde ke zpřesnění specifikací

Harmonogram

Rádi bychom v září na konferenci Knihovny současnosti předvedli odborné veřejnosti první výsledky

Veřejnost by se k portálu mohla dostat na začátku příštího roku



KNIHOVNY.CZ

Děkuji za pozornost

cpk@mzk.cz