

Projekt INTERPI

Jana Šubová

Cosmotron Bohemia, s.r.o.

subova@cosmotron.cz

Marie Balíková

Národní knihovna ČR

marie.balikova@nkp.cz

Miroslav Kunt

Národní archiv

miroslav.kunt@nacr.cz

Nadežda Andrejčíková

Cosmotron Bohemia, s.r.o.

andrejickova@cosmotron.cz

INFORUM 2016: 22. ročník konference o profesionálních informačních zdrojích
Praha, 24. - 25. 5. 2016

Abstrakt

Příspěvek přináší informace o projektu Interoperabilita v paměťových institucích (INTERPI) DF11P01OVV023 Programu aplikovaného výzkumu a vývoje národní a kulturní identity (NAKI) financovaném Ministerstvem kultury ČR.

Fondy a sbírky paměťových institucí tvoří jádro národního kulturního, industriálního a přírodního dědictví. Cílem projektu INTERPI je tvorba společné ontologie a znalostního modelu odpovídajícího potřebám všech paměťových institucí obohaceného o potřebné sémantické informace umožňující i strojové zpracování dat. Projekt je zaměřen především na ty entity, které tvoří základ pro propojení objektů z fondů a sbírek paměťových institucí. Přináší také nové paradigma zpracování dat na základě objektového přístupu, které se zaměřuje na zpracování entit (tříd) a komplexních vztahů mezi nimi, jehož cílem je zabezpečení sémantické interoperability.

Reaguje na potřebu změnit přístup k vlastní koncepci zpřístupnění informací s ohledem na sémantický web, odvážně a kriticky zhodnotit stávající standardy, pravidla a postupy a hledat nové možnosti. Projekt INTERPI je systematickým pokusem o přípravu takové cesty a to napříč knihovnami, muzei, galeriemi i archivy. Přínosem projektu je, že výrazným způsobem přispěl k otevření diskuse mezi různými komunitami z paměťových institucí.

Úvod

V roce 2010 byl v Programu aplikovaného výzkumu a vývoje národní a kulturní identity (NAKI) financovaném Ministerstvem kultury ČR podán projekt Interoperabilita v paměťových institucích (INTERPI). Na konci roku 2010 byl projekt přijat a 1. 2. 2011 bylo zahájeno jeho řešení. Projekt byl naplánován jako pětiletý, s plánovaným ukončením 31. 12. 2015. O průběhu řešení projektu jsme pravidelně informovali na odborných akcích, přičemž jsme se zaměřili na výsledky jednotlivých etap řešení. Cílem tohoto příspěvku je shrnout především praktické výsledky a zkušenosti z řešení projektu a načrtnout jeho další „život“ s ohledem na požadavky odborných komunit.

Projekt INTERPI je specifický účastí dvou institucí v gesci dvou ministerstev (Národní knihovna ČR – Ministerstvo kultury a Národní archiv – Ministerstvo vnitra). Takto postavený řešitelský tým znamenal užší spolupráci dvou odborných komunit a současně vyšší administrativní náročnost.

První část příspěvku přináší informace o konkrétních výsledcích projektu v podobě metodik, softwarového řešení atd. Mezi zásadní výsledky projektu (v abstraktní rovině) patří dosažení

konsenzu v požadavcích různých paměťových institucí (archivů, knihoven, muzeí a galerií), proto je druhá část příspěvku zaměřena právě na vztah projektu k jednotlivým typům paměťových institucí a k jeho dalšímu předpokládanému rozvoji. Třetí část příspěvku je věnována projektu z pohledu „běžného“ uživatele – tj. uživatele, který využívá informace ze znalostní báze na přístup ke kulturnímu, industriálnímu a přírodnímu dědictví ve správě paměťových institucí.

Výsledky projektu

Jak již bylo zmíněno v úvodu, výsledky projektu jsme průběžně prezentovali na odborných akcích, proto se v tomto příspěvku zmíníme především o těch, které jsou podle našeho mínění důležité pro další využití, provoz a rozvoj projektu. Jde o výsledky aplikovaného výzkumu, tj. metodiky, software a tzv. funkční vzorek.

Metodika tvorby znalostního modelu

Tvorbě metodiky předcházely činnosti:

- výčet typů entit, které jsou předmětem zájmu paměťových institucí - představoval první střet různé terminologie a praxe paměťových institucí, výsledkem bylo částečné opuštění knihovnické terminologie a ustálení terminologie, která je pro všechny komunity dostatečně významově neutrální;
- analýza standardů souvisejících s činnostmi paměťových institucí, při které jsme se zaměřili na definice a vlastnosti entit;
- specifikace požadavků na rozsah a strukturu informací potřebných pro popis entit – výsledkem byl rozsáhlý seznam různých vlastností, které by znalostní databáze měla zachytit;
- generalizace vlastností a vytvoření konceptuálního modelu, který by pokrýval rozsah požadovaných informací, byl dostatečně univerzální a otevřený a minimalizoval duplicitu prvků;
- vymezení tříd entit na základě společných vlastností.

Metodika tvorby znalostního modelu tedy vychází z obecného konceptuálního modelu znalostní databáze INTERPI a vymezení tříd entit. Jejím cílem není opakovat pravidla využívání jednotlivými komunitami, ale hledat kompromis mezi jejich rozdílnými požadavky.

Metodika mapování a harmonizace rejstříků, číselníků a řízených slovníků aplikovaných v paměťových institucích

Tato metodika vznikla z potřeby řešit otázky propojování různých typů zdrojů termínů. Předpokládáme, že INTERPI využijí i odborné komunity, které nemají k dispozici nástroje na budování slovníků, proto metodika obsahuje doporučení pro autorizaci nových termínů a/nebo propojení s existujícími termíny.

Rozhraní pro zpracování entit

Zpracovatelé mají k dispozici webové rozhraní pro úpravu a vkládání záznamů všech typů entit. Rozhraní využívají i supervizoři ke kontrole vytvořených záznamů a odsouhlasení změn. Funkce rozhraní jsme již více prezentovali na konferencích Archivy, knihovny a muzea v digitálním světě.

Software pro zpřístupnění a prezentaci znalostní databáze INTERPI

Záznamy entit jsou prezentovány ve webovém rozhraní pro zpřístupnění databáze. Rozhraní je určeno především běžným uživatelům, umožňuje vyhledání a zobrazení informací o entitě. Umožňuje nalezené záznamy odeslat mailem, uložit v uživatelsky přístupném formátu nebo exportovat ve strojem čitelných formátech (zkrácené XML, kompletní XML, rozšířené MARC21 XML).

Funkční vzorek

Funkční vzorek představuje soubor technologických prvků, metodických pokynů a vymezení uživatelských skupin a rolí, které vytvářejí řešení pro provoz databáze INTERPI. Mezi technologické prvky patří datové úložiště a databáze, webový a aplikační server a technologie pro zpřístupnění dat (webové služby, Z39.50 server) a konkrétní softwarové řešení pro vytváření, provoz a prezentaci znalostní databáze INTERPI. Uživatelské skupiny a role určují, kdo a za jakých podmínek má přístup k jednotlivým částem řešení. Metodické pokyny (v podobě výše zmíněných metodik) představují rámec pro vytváření znalostní databáze.

INTERPI ve vztahu k paměťovým institucím

INTERPI a archivy

Projekt INTERPI vzbudil odezvu především v prostředí archivů. Souběžně s realizací projektu INTERPI se rozvinulo více aktivit na přípravě metodických a metodologických postupů v oblasti zpracování archivních dokumentů a jejich zavedení v archivech. Aktivity navazovaly na předešlé snahy o přiblížení standardů pro archivní popis v podobě překladů dokumentů ISAD (G) a ISAAR (CPF). V této souvislosti byl vytvořen návrh nových *Základních pravidel pro zpracování archivních materiálů*, jehož koncepce vychází z mezinárodních archivních standardů. Práce na Základních pravidlech pro zpracování archiválií byly v případě kapitol „popis původců“ a „tvorba názvů a jmen“ koordinovány a konzultovány i v rámci projektu INTERPI. Konzultace přispěly k definici entit (nejen základních) a také k zamyšlení nad udržitelností či neudržitelností některých metodických postulátů ať již v oblasti archivnictví nebo knihovnictví.

Taktéž při přípravě *novely zákona 499/2004 Sb., o archivnictví a spisové službě a novele vyhlášky 645/2004 Sb., kterou se provádí zákon o archivnictví a spisové službě* byly práce navázány na projekt INTERPI. Do zákona byla při přípravě novely včleněna ustanovení o evidenci původců, z nichž nejvýznamnější je § 18c odst. 3, který kodifikuje povinný popis původců v elektronické podobě jednotlivými archivy. „Popis a evidence původců“ jsou postaveny na roveň (paralelně vedle ní) dosavadní evidenci Národního archivního dědictví. Při aplikaci tohoto ustanovení je zamýšleno užívání databáze INTERPI, pro niž komunikaci bude zprostředkovávat národní archivní portál. Novela vyhlášky upravuje například definici korporací v souladu s jejím vymezením pro INTERPI, které opouští dosavadní koncept pouze právnických osob.

Praktické implementace výsledků projektu INTERPI v archivech představují *propojení s informačními systémy*:

- Aplikace pro popis archiválií Heritage – ve fázi testování možností propojení.
- Software ELZA – elektronické zpracování archiválií vyvíjený na základě veřejné soutěže Technologické agentury ČR „Vývoj software pro popis archiválií na základě Základních pravidel pro zpracování archiválií metodou PCP“.
- Národní archivní portál jako součást národního digitálního archivu.
- Popis původců v software pro evidenci národního archivního dědictví PevA.

Z vyjmenovaných systémů je vhodné zdůraznit Národní archivní portál a software PevA. *Národní archivní portál* je součástí národního digitálního archivu a zákonem deklarován jako informační systém veřejné správy. Zadání jeho řešení bylo veřejnou soutěží zadáno z prostředků Národního archivu, přičemž využití záznamu z INTERPI je předepsáno v případě názvu archivu (všechny archivy v ČR) a původce. Modul pro evidenci původců je přímo určen pro komunikaci s INTERPI, přičemž udržuje lokální bázi dat.

V aplikaci pro evidenci Národního archivního dědictví (PevA) vyvíjené a distribuované všem institucím, které vedou evidenci archiválií, Ministerstvem vnitra, byla doplněna samostatná evidence původců, která vychází z metodiky znalostního modelu INTERPI a je na znalostní databázi INTERPI napojena.

The screenshot shows the National Archival Portal search interface. At the top, there is a navigation bar with the logo and text 'Národní archiv', 'Národní archivní portál', and search options. Below this, there are tabs for 'Vyhledávání v archívech', 'Archivy', 'Vyhledávání ve jménech', and 'Původci'. The main search area is divided into 'Volby vyhledávání' (Search filters) and 'Hledaný řetězec' (Search string). The search string is 'ministerstvo financí'. The search results section shows 'Počet výsledků : 6' and a table with columns 'Jméno', 'Popis', 'Kategorie', and 'Akce'.

Jméno	Popis	Kategorie	Akce
Česko (Československo) Ministerstvo financí		korporace	Q
Československo. Federální ministerstvo financí		korporace	Q
Československo. Ministerstvo financí		korporace	Q
Česko. Ministerstvo financí		korporace	Q
Česko. Ministerstvo financí. Centrum pro zahraniční pomoc		korporace	Q
Slovensko. Ministerstvo financí		korporace	Q

Obrázek 1: Ukázka vyhledání původců s využitím databáze INTERPI v archivním portálu

Na propojení archivních informačních systému s INTERPI se využívají webové služby. Umožňují vyhledat v databázi INTERPI záznamy podle vymezených kritérií nebo pomocí jejich kombinací. Výsledkem dotazu jsou vyhovující záznamy v proprietárním XML formátu se základními údaji, které umožňují jejich spolehlivou identifikaci, rozlišení od jiných entit a posouzení relevance pro další využití. Jako základní údaje byly vybrány preferovaná i variantní označení, stručná charakteristika, hierarchická struktura geografické entity, identifikační čísla – IČO, kód_cz, číslo Českého statistického úřadu, číslo Slovenského statistického úřadu, identifikátor INTERPI. S využitím jiné webové služby je možné získat kompletní záznam entity taktéž v proprietárním XML formátu se všemi souvislostmi – vazbami a událostmi. Nenajde-li se záznam relevantní pro další využití, je možné zase pomocí webové služby záznam v databázi INTERPI založit. Pro založení je doporučována struktura základních údajů (viz výše) a data jsou předávána ve stejném proprietárním XML formátu.

INTERPI a knihovny

Pro knihovny není koncept budování a využívání centrální databáze (např. Národní autority) nijak nový. Také standardizace zpracování má dlouhou tradici. Ve vztahu k projektu INTERPI je však pro knihovny novinkou rozsah údajů o entitách, který je možné zaznamenávat a také větší fragmentace údajů a důraz na vytváření vztahů mezi entitami.

Pro knihovny jsou jistě zásadní dvě otázky: *jaký je vztah databáze INTERPI k Národním autoritám a jak mohou knihovny využívat v budoucnu databázi INTERPI.*

Na první otázku (vztah INTERPI a Národních autorit) jsme odpovídali v průběhu řešení projektu a podrobnější informace lze najít především v Metodice mapování a harmonizace rejstříků, číselníků a řízených slovníků aplikovaných v paměťových institucích. *Národní autority jsou jedním z terminologických zdrojů pro INTERPI* – ne všechny entity, které mohou vzniknout v INTERPI musí být zařazeny do jednoho ze souborů národních autorit a také je přirozené, že ne všechny zpracovávané údaje o entitách lze prezentovat ve formátu souboru národních autorit (MARC21). Entity, které v INTERPI vznikají na základě záznamů národních autorit, mají uchované identifikační číslo národní autority a také typ autority. Jsme si vědomi, že automatická synchronizace záznamů entit podle změn v souborech národních autorit, může být v databázi INTERPI problematická (údaje ze záznamů Národních autorit jsou transformovány do struktury pro INTERPI a doplňovány zpracovateli), proto se předpokládá pouze pro záznamy, které byly z Národních autorit přebrány a nebyly doposud upraveny.

Knihovny mohou využívat databázi INTERPI prostřednictvím protokolu Z39.50 stejně jako jiné zdroje a soubory autorit. Pro prezentaci dat protokolem Z39.50 jsou záznamy entit transformovány do formátu MARC21. K dispozici je základní záznam pro potřeby souborů autorit (v rozsahu stanoveném metodikami a doporučeními Národní knihovny ČR).

```

SYS      n000109849
LDR      -----nz---a22-----n--4500
003      INTERPI
005      20151201174338.9
008      020513|n|acnnaabn-----n-a|a-----
040      $$aABG001$$bce$$dABA001$$dINTERPI
046      $$f1814$$g1902
100 1    $$aSchmoranz, František$$d1814-1902
370      $$aBělá, Liberk, Česko$$bSlatiňany, Česko$$cČesko
          $$fChrudim, Česko
372      $$aarchitektura$$astavitelství
374      $$aarchitekti$$akonzervátoři
375      $$amuž
377      $$acze$$ager
400 1    $$aSchmoranz, Franz$$d1814-1902
400 1    $$ašmoranc, František$$d1814-1902
400 1    $$aSchmoranz, František$$d1814-1902$$cst.
670      $$aI. Šulc: Rod Schmoranzů ze Slatiňan. Slatiňany 2002
670      $$aStátní okresní archiv Chrudim, fond Rodinný archiv
          Schmoranzů
670      $$aStátní oblastní archiv v Zámrsku, Sběrka matrik
          východočeského kraje, sign. 3-1, Bělá - matrika narozených
          1784 až 1833
678 0    $$aStavitel ve Slatiňanech, městský architekt v Chrudimi,
          krajský konzervátor památek, často zaměňován se svým synem
          Františkem (1845-1892)

```

Obrázek 2: Příklad zkráceného MARC21 formátu záznamy entity z databáze INTERPI

Druhá podoba záznamu entity ze znalostní databáze INTERPI ve formátu MARC21 představuje rozšířený záznam, který využívá možnosti formátu MARC21 a obsahuje vazby mezi entitami a další údaje nad rámec metodik a doporučení Národní knihovny. Rozšířený záznam ve formátu MARC21 XML je k dispozici jako jeden z exportních formátů ve webovém rozhraní pro zpřístupnění a prezentaci entit. Podle zájmu a odezvy může být rozšířený MARC21 formát alternativním formátem pro webové služby.

```

SYS      n000109849
LDR      -----nz--a22-----n--4500
003      INTERPI
005      20151201174338.9
008      020513|n|acnnaabn-----n-a|a-----
040      $$aABG001$$bcze$$dABA001$$dINTERPI
046      $$f18141230$$g19020404
100 1    $$aSchmoranz, František$$d1814-1902
370      $$aBělá, Liberk, Česko$$0ge494224$$bSlatiňany, Česko
          $$0ge130838$$cČesko$$0ge128065$$fChrudim, Česko$$0ge128952
372      $$aarchitektura$$0ph118586$$astavitelství$$0ph126091
374      $$aarchitekti$$0ph118582$$akonzervátoři$$0ph565265
375      $$amuž
377      $$acze$$ager
400 1    $$aSchmoranz, Franz$$d1814-1902
400 1    $$ašmoranc, František$$d1814-1902
400 1    $$aSchmoranz, František$$d1814-1902$$cst.
500 1    $$aSchmoranz, František$$d1845-1892$$isyn$$wi$$7jo2001100097
670      $$aI. Šulc: Rod Schmoranzů ze Slatiňan. Slatiňany 2002
670      $$aStátní okresní archiv Chrudim, fond Rodinný archiv
          Schmoranzů
670      $$aStátní oblastní archiv v Zámrsku, Sbíрка matrik
          východočeského kraje, sign. 3-1, Bělá - matrika narozených
          1784 až 1833
678 0    $$aStavitel ve Slatiňanech, městský architekt v Chrudimi,
          krajský konzervátor památek, často zaměňován se svým synem
          Františkem (1845-1892)

```

Obrázek 3: Příklad rozšířeného formátu MARC21 (pro čitelnost prezentováno jako řádkový MARC), zvýrazněné části porovnej s obrázkem 2

Při zadávání projektu INTERPI jsme počítali s tím, že v průběhu realizace dojde k zavedení pravidel RDA do knihoven, které poskytnou větší prostor pro strukturalizovaný zápis informací o autoritě a zachycení více vztahů. Stalo se tak sice se zpožděním oproti plánované implementaci, ale důležité je, že pravidla RDA jsou pro autority zavedena a zmíněné dvě verze záznamu nejsou díky nim tak rozdílné, jak by tomu bylo v minulosti.

Využití databáze INTERPI se předpokládá především v knihovnách, které budují databáze regionálních autorit a již v této době vytvářejí obohacené záznamy. Databáze INTERPI pro ně může být zajímavým zdrojem především za předpokladu, že bude obsahovat informace doplňované archivy.

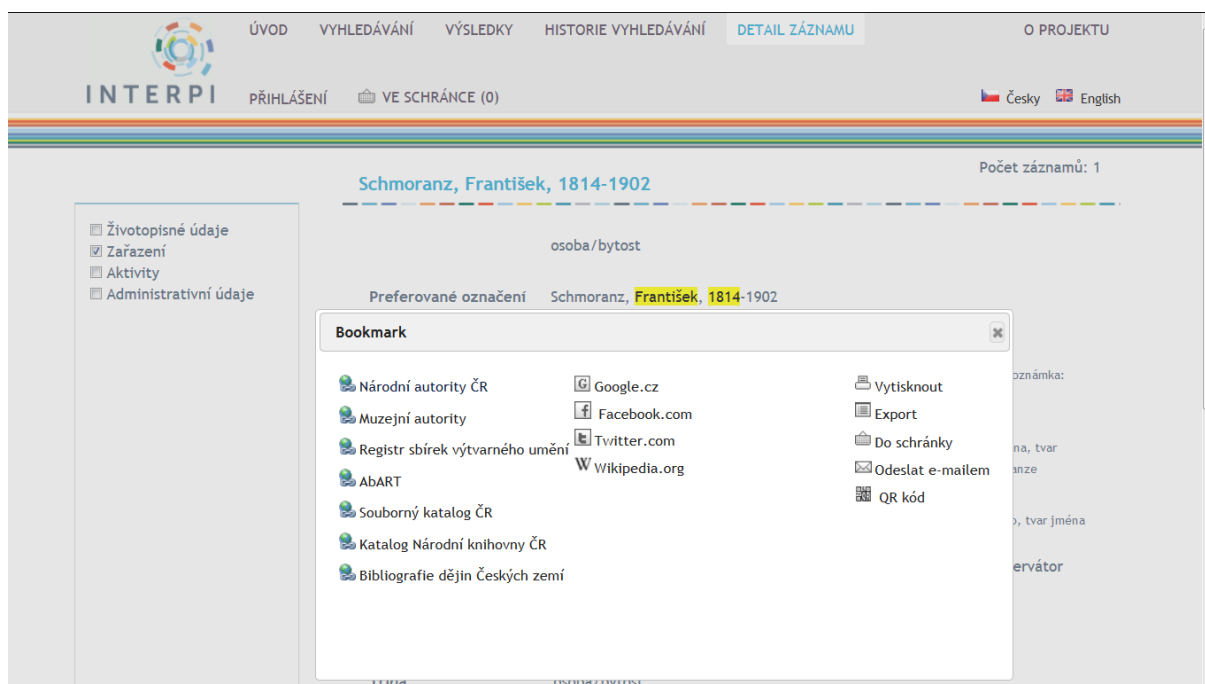
INTERPI a muzea a galerie

V oblasti muzeí a galerií je důležité vymezit vztah projektu INTERPI k výsledkům projektu Muzejních autorit. INTERPI na projekt Muzejních autorit navazuje a respektuje jeho závěry v oblasti personálních autorit. Struktura dat entit osoba/bytost umožňuje zachytit všechny vztahy, které jsou vymezeny v Muzejních autoritách. Muzejní autority jsou pro INTERPI významným terminologickým zdrojem. Záznamy, které jsou z Muzejních autorit přebrány, si uchovávají identifikátor a typ autority.

INTERPI ve vztahu k běžným uživatelům

Zadání projektu INTERPI se na běžného uživatele zaměřilo především v oblasti jednotného zpřístupnění kulturního, přírodního a industriálního dědictví. Pokud se v archivech rozšíří využívání entit INTERPI na identifikaci původců, bude databáze INTERPI dalším přístupovým bodem k záznamům například v národním archivním portálu s přesahem do oblasti veřejné správy.

Další rozvoj INTERPI bude směřovat k propojení dat s jinými informačními systémy na základě identifikátorů entit.



Obrázek 4: Příklad odkazů do dalších zdrojů z rozhraní pro prezentaci entit

Závěr – budoucnost projektu

Jak již bylo naznačeno v předchozích částech příspěvku, budoucnost projektu směřuje k využití výsledků projektu jednotlivými komunitami i běžnými uživateli. Prostor pro další rozvoj vidíme především v iniciativách e-Culture, které jsou zaměřené právě na zpřístupnění kulturního dědictví a v aktivitě zúčastněných subjektů – všech paměťových institucí.