

ZPRÁVA O ČINNOSTI NÁRODNÍHO ARCHIVU V PROJEKTU INTERPI ZA ROK 2015

INTERPI – Interoperabilita v paměťových institucích

Program aplikovaného výzkumu a vývoje národní a
kulturní identity (NAKI)
(DF11P010VV023)

Sestavil:
Miroslav Kunt



INTERPI

30. 10. 2015 | verze 1.0

Obsah

1	Spuštění a testování webových služeb	2
2	Implementace INTERPI.....	6
2.1	Národní archivní portál	6
2.2	Popis původců v software pro evidenci národního archivního dědictví PevA.....	8
3	Pořizování záznamů o entitách.....	9
4	Příprava metodických a metodologických postupů, jejich zavedení v archivech	10

1 Spuštění a testování webových služeb

Spuštění poloprovozu databáze INTERPI a rozpracovanost navazujících řešení v archivnictví vyžadovaly součinnost mezi Národním archivem a dodavatelem služby (Cosmotron Bohemia s.r.o.) vysoutěžené v roce 2014 s faktickou opcí na dokončení v roce 2015 (viz komentář k rozpočtu). Jako zcela zásadní se ukázala otázka zprovoznění webových služeb, na jejichž základě budou s INTERPI komunikovat další systémy. Základní náčrt webových služeb byl podán v prvním roce řešení projektu.¹ Práce navázaly na návrh webových služeb z roku 2014 (viz předchozí průběžná zpráva). V roce 2015 se již jednalo o praktické testování a shromažďování připomínek k úpravám tak, aby mohl být v závěru roku otestován poloprovoz i po této stránce a zároveň bylo možné praktické nasazení INTERPI do provozu archivů a národního archivního portálu jako součásti národního digitálního archivu.

Celé řešení webových služeb umožňuje použití obou způsobů pro výměnu zpráv mezi klientem a serverem - SOAP (Simple Object Access Protocol) a REST (Representational State Transfer). Komunikace je zatím uskutečňována prostřednictvím protokolu HTTP narozdíl od prostředí knihoven, kde je tradičně využíván protokol Z 39.50 (ISO 23950). Pro dotazování je využit jazyk PQF (Prefix Query Format), pro jehož použití byla vyhotovena dodavatelem služby specifikace.

PQF - přehled implementovaných vyhledávacích kritérií:²

- **2050** - stav záznamu - obsahuje kód pro stav záznamu vzhledem na jeho životní cyklus:
 - **nr** - rozpracovaný návrh
 - **nd** - dokončený návrh
 - **rr** - rozpracovaná revize
 - **rd** - dokončená revize
 - **sz** - schválený záznam
 - **zz** - zrušený záznam
 - **ir** - bez revize INTERPI (záznamy získané importem z jiných systémů)
- **2051** - třída - obsahuje jednoznačný kód pro označení třídy entit:
 - **o** - třída osoba/bytost
 - **r** - třída rod/rodina
 - **k** - třída korporace
 - **g** - třída geografický objekt
 - **u** - třída událost
 - **d** - třída dílo/výtvor
 - **p** - třída obecný pojem
- **2054** - preferované označení entity - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla a bez ohledu na zařazení do třídy - při použití jako kompletního termínu obsahuje hlavní, vedlejší část označení a doplňky

¹ Šubová, Jana, Návrh webových služeb pro komunikaci s databází INTERPI, 2011, 15 s.

² Šubová, Jana, Použití dotazovacího jazyku PQF v INTERPI, 2015, 5 s.

- **2055** - všechny označení entity - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení a zařazení do třídy - použití jako kompletního termínu viz atribut 2054
- **2056** - preferované označení entit třídy osoba/bytost - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla označení - použití jako kompletního termínu viz atribut 2054
- **2057** - všechny označení entity třídy osoba/bytost - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení - použití jako kompletního termínu viz atribut 2054
- **2058** - preferované označení entit třídy rod/rodina - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla označení - použití jako kompletního termínu viz atribut 2054
- **2059** - všechny označení entity třídy rod/rodina - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení - použití jako kompletního termínu viz atribut 2054
- **2060** - preferované označení entit třídy korporace - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla označení - použití jako kompletního termínu viz atribut 2054
- **2061** - všechny označení entity třídy korporace - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení - použití jako kompletního termínu viz atribut 2054
- **2062** - preferované označení entit třídy událost - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla označení - použití jako kompletního termínu viz atribut 2054
- **2063** - všechny označení entity třídy událost - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení - použití jako kompletního termínu viz atribut 2054
- **2064** - preferované označení entit třídy geografický objekt - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla označení - použití jako kompletního termínu viz atribut 2054
- **2065** - všechny označení entity třídy geografický objekt - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení - použití jako kompletního termínu viz atribut 2054
- **2066** - preferované označení entit třídy dílo/výtvar - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla označení - použití jako kompletního termínu viz atribut 2054
- **2067** - všechny označení entity třídy dílo/výtvar - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení - použití jako kompletního termínu viz atribut 2054
- **2068** - preferované označení entit třídy obecný pojem - obsahuje všechny preferované označení entity bez ohledu na aplikovaná pravidla označení - použití jako kompletního termínu viz atribut 2054
- **2069** - všechny označení entity třídy obecný pojem - obsahuje všechny označení entity bez ohledu na aplikovaná pravidla, jazyk nebo preferované/variantní označení - použití jako kompletního termínu viz atribut 2054

- **2008** - obecný doplněk - obsahuje obecný doplněk v označení entit, kde je obecný doplněk využíván
- **2009** - chronologický doplněk - obsahuje obecný doplněk v označení entit, kde je chronologický doplněk využíván
- **2010** - geografický doplněk - obsahuje obecný doplněk v označení entit, kde je geografický doplněk využíván
- **2011** - příslušnost k zemi původu - obsahuje hlavní část označení (preferovaných i variantních) entity geografický objekt, která je využita pro vyjádření příslušnosti k zemi původu
- **2012** - obor působnosti - obsahuje hlavní část označení (preferovaných i variantních) entity obecný pojem, která je využita pro vyjádření oboru působnosti
- **2013** - kategorie - obsahuje hlavní část označení (preferovaných i variantních) entity obecný pojem, která je využita pro vyjádření kategorie entity
- **2014** - jazyková oblast - obsahuje kód jazykové oblasti entity
- **2015** - datum vzniku - obsahuje datum vzniku entity
- **2016** - datum zániku - obsahuje datum zániku entity
- **2017** - místo vzniku - obsahuje hlavní část označení (preferovaných i variantních) entity geografický objekt, která je využita pro vyjádření místa vzniku entity
- **2018** - místo zániku - obsahuje hlavní část označení (preferovaných i variantních) entity geografický objekt, která je využita pro vyjádření místa zániku entity
- **2019** - sídlo - obsahuje hlavní část označení (preferovaných i variantních) entity geografický objekt, která je využita pro vyjádření sídla entity
- **2020** - geografická působnost - obsahuje hlavní část označení (preferovaných i variantních) entity geografický objekt, která je využita pro vyjádření geografické působnosti entity
- **1016** - vše - obsahuje údaje ze všech částí záznamu
- **12** - INTERPI identifikační číslo záznamu

Zejména tato vyhledávací kritéria se ukázala klíčová pro definování způsobu dotazování. Druhou podstatnou otázkou jsou vlastní data, která jsou webovou službou předávána prostřednictvím XML definovaným XML schématem (XSD). Toto schéma je proprietární tak, aby umožňovalo vysokou flexibilitu při implementaci INTERPI v různých aplikacích v odlišných paměťových institucích, kde nestačí jen základní informace o entitě.

V zásadě jsou definovány dvě XML schémata:

- pro seznam záznamů vyhledaných na základě dotazu,
- pro jeden záznam.

Prostřednictvím XML pro jeden záznam nejsou zasílána kompletně všechna data - ta jsou k dispozici na vlastním webovém rozhraní INTERPI a není důvod toto rozhraní dublovat. Každopádně všechny údaje potřebné pro rejstříky i popis původců v archivních pomůckách v XML obsaženy jsou. Tento rozsah bude nepochybně vyhovovat i ostatním paměťovým institucím, které se komunikaci prostřednictvím webových služeb rozhodnou využívat. Nedostatek v podobě chybějících záznamů napojených entit, které v XML byly reprezentovány pouze ID INTERPI (bylo by nutné pro všechny napojené entity dotaz opakovat a následně by byly zaslány jejich úplné záznamy) je odstraňován.

Seznam záznamů vyhledaných na základě dotazu musí obsahovat pouze základní údaje, na jejichž základě může uživatel rozhodnout - vybrat konkrétní odpovídající entitu. To nastolilo otázku, které údaje jsou z pohledu uživatele těmi podstatnými. Je jím především preferované jméno - ale preferované podle jakých pravidel? Proto bylo rozhodnuto doplnit do dotazu identifikaci typu paměťové instituce (archiv - knihovna - muzeum) a podle ní vybrat preferované jméno. V řadě případů však dochází k tomu, že uživateli nebude postačovat jen preferované jméno, ale bude potřebovat srovnat i jména variantní. Např. při zadání dotazu prostřednictvím nepreferovaného cizojazyčného názvu může být vrácena sada záznamů, z jejichž pouze preferovaných jmen (ani dalších údajů, např. stručné charakteristiky) nebude možné nebo snadné vybrat. Teprve srovnání nepreferovaných jmen může vést k cíli. Stejně tak budou potřeba dalších informací pro rozhodnutí - zejména stručné charakteristiky v případě osob, dat vzniku/zániku v případě korporací apod. Nabízí se tedy vracet na dotaz sadu záznamů obsahující všechny tyto údaje. V takovém případě nebude ani nutné pro jednoduché použití v rejstřících apod. požadovat úplná data u vybraného záznamu.

2 Implementace INTERPI

V různém stádiu rozpracovanosti je implementace INTERPI do čtyřech řešení, na jejichž přípravě se Národní archiv podílel a podílí:

1. Aplikace pro popis archiválií Heritage (testování).
2. 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“.
3. Národní archivní portál jako součást národního digitálního archivu.
4. Popis původců v software pro evidenci národního archivního dědictví PevA

Na úspěchu poloprovozu databáze INTERPI včetně webových služeb závisí i funkčnost výše zmíněných projektů, přičemž v případě 2 a 3 jde o funkčnost nepostradatelnou. Proto také podněty z řešení těchto aplikací jsou zpracovávány a je tak ještě před zpuštěním poloprovozu zajištěna zpětná vazba budoucích uživatelů.

2.1 Národní archivní portál

Národní archivní portál je součástí národního digitálního archivu a zákonem deklarován jako informační systém veřejné správy. Z pohledu zákona o kybernetické bezpečnosti jde o kritický informační systém. 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. V případě původců se jedná o všechny původce archiválií v digitální podobě a původce, kteří předkládají návrh na výběr archiválií - z hlediska zákona všichni, kdo vedou spisovou službu v elektronických systémech spisové služby (celá veřejná správa v ČR). Modul pro evidenci původců je přímo určen pro komunikaci s INTERPI, přičemž udržuje lokální bázi dat. Zadání bylo upraveno (jedná se o modifikaci požadavků uvedených v minulé průběžné zprávě):

1. **Autentizace a autorizace:** Modul vyžádá prostřednictvím webové služby od modulu pro administraci uživatelů autorizaci autentizovaného uživatele nebo služby před provedením činnosti.
2. **Zaznamenání operací:** Údaje o provedených operacích vykonaných modulem a uživateli předává modul webovou službou modulu pro logování a vyhodnocování.
3. **Zdroj dat:** Primární evidencí původců je databáze INTERPI (www.interpi.cz). Ve smyslu tohoto projektu se jedná o korporace, rody, osoby a události - dočasné korporace. Synchronizace s vyhledávací databází modulu probíhá automatizovaně nebo manuálně na základě pokynu oprávněného uživatele v roli Správce, který může stáhnout pro modul seznam změněných záznamů od poslední synchronizace (vystavuje INTERPI). Modul porovná na základě ID INTERPI seznam změněných záznamů s vlastní databází a vytvoří seznam změněných záznamů. Odešle seznam změněných záznamů (jejich ID INTERPI). Databáze INTERPI vrátí na základě ID INTERPI úplné záznamy (u kterých došlo ke změnám). Modul provede změny ve svých záznamech tím, že je přepíše získanými hodnotami z INTERPI.

Modul poskytuje webovou službu modulu pro centrální rozhraní uživatelskou verzi databáze tvořenou vybranými záznamy.

Při získávání záznamu oprávněný uživatel vyhledává původce pomocí zadání dotazu (jednoduchého nebo strukturovaného). Komunikace se děje ve dvou stupních. Nejprve se prohledává vlastní databáze původců v modulu (dotaz 1/ odpověď 1), teprve konkrétní úplný záznam se zobrazí stažením z INTERPI (dotaz 2/ odpověď 2).

Modul zodpoví dotaz zasláný uživatelem prostřednictvím modulu pro centrální rozhraní a to optimalizovaným způsobem pro účelné zobrazení výsledku a na vyžádání poskytuje další výsledky pro zobrazení:

1. Modul vyhledává dotaz ve vlastní databázi a zobrazí seznam.
2. V případě požadavku na konkrétní úplný záznam (podrobnosti) modul odešle na základě výběru ze seznamu ad 1) dotaz prostřednictvím webové služby do databáze INTERPI s ID záznamu INTERPI; je možné vyznačit víc záznamů v seznamu - maximálně 100 najednou a maximálně 300 za den pro jednoho uživatele. Databáze INTERPI vrátí jeden nebo více záznamů odpovídajících dotazu (ID INTERPI) - v úplném tvaru (všechny údaje v definované podobě v XML). Modul zobrazí jednotlivé zasláné záznamy v úplné podobě (seznam i konkrétní záznam je možné stáhnout do textového souboru nebo vytisknout). Obdobně se postupuje při zaslání dotazu k vazbám mezi původci a jinými entitami (související entity) - pracuje se s odkazem na entitu přítomnou v úplném záznamu.

Jestliže záznam o původci neexistuje v databázi modulu, tento sestaví dotaz a odešle ho prostřednictvím webové služby do databáze INTERPI. INTERPI odpoví v podobě Ano/Ne (1/0) - tedy zda danému požadavku odpovídá nějaký záznam. Podle toho dojde k 1) zaslání přehledu záznamů, 2) iniciaci vytvoření nového záznamu:

1. Databáze INTERPI vrátí jeden nebo více záznamů (v XML souboru) odpovídajících dotazu od pozice x v rozsahu y záznamů (např. záznamy 1-10) - pouze základní údaje (zejména preferovaná forma jména včetně doplňku, datum a místo narození/úmrť, vzniku/zániku, začátku/konce, stručná charakteristika - podrobnosti v příloze 10 Základních pravidel pro zpracování archiválií, dostupných na <http://www.mvcr.cz/soubor/zakladni-pravidla-pro-zpracovani-archivalii-2013-pdf.aspx>). Pokud je v syntaxi dotazu definován parametr pro setřídění, bude výsledek setříděn podle tohoto parametru. Modul zobrazí ve svém zobrazovacím formátu seznam zasláných odpovědí na dotaz; možné následně zobrazit i celé záznamy dle postupu uvedeného výše.
2. Modul na základě autorizace ověří oprávnění uživatele a provede jeho přihlášení prostřednictvím příslušné skupiny uživatelů k INTERPI (např. Archiv XY). Prostřednictvím tlačítka „Vytvořit záznam“ modulu se otevře rozhraní INTERPI - volá se URL INTERPI a provede se přihlášení agregovaného uživatele „Archiv XY“ s parametrem konkrétního uživatele autentizovaného a autorizovaného modulem.

Po dokončení editace záznamu v INTERPI jeho uložením dojde ke stažení části údajů (v rozsahu dle §12b vyhl. č. 645/2004 Sb.) o původci do databáze modulu; ostatní údaje jsou vedeny v

databázi INTERPI, stejně jako vazby mezi entitami. Na základě identifikace uživatele „Archiv XY“ bude INTERPI volat webovou službu modulu pro uložení těchto vybraných dat k modulu.

V případě nedostatečného záznamu nebo jiné potřeby editovat existující záznam v databázi INTERPI se postupuje shodně jako při tvorbě nového záznamu, ale použije se modulem zobrazené tlačítko „Upravit záznam“.

Po dokončení editace záznamu v INTERPI jeho uložení dojde ke stažení části údajů (v rozsahu dle §12b vyhl. č. 645/2004 Sb.) o původci do databáze modulu, kde nahradí dosavadní. Změna se zaznamená do transakčního protokolu (nikoli pouze druh změny a pole, ale původní znění). Ostatní údaje jsou vedeny v databázi INTERPI, stejně jako vazby mezi entitami.

2.2 Popis původců v software pro evidenci národního archivního dědictví PevA

V aplikaci pro evidenci Národního archivního dědictví vyvíjené a distribuované všem institucím, které vedou evidenci archiválií, Ministerstvem vnitra, byla doplněna samostatná evidence původců, která vychází z metodiky znalostního modelu INTERPI. Použití je pro nově vytvořené archivní soubory (po 1. červenci 2015) povinné, v Národním archivu byly transformovány a doplněny záznamy i zpětně - podrobnosti jsou uvedeny v monografii o projektu INTERPI (autor Mgr. Rudolf Kad'orek). Lze konstatovat, že metodika vypracovaná v projektu INTERPI je funkční, ale bude nutné na základě připomínek uživatelů znovu některé věci posoudit (příkladem je absence zdravotnických zařízení v podtřídách třídy korporace a jejich doplnění).

3 Pořizování záznamů o entitách

Práce na záznamech o entitách, resp. shromažďování dat, pokračovalo i v roce 2015, i když ne již takovým tempem. Příčinou bylo jednak relativně pozdní zprovoznění vlastní „ostré“ databáze (obtíže s importem dat z národních autorit), jednak se pozornost přesunula k složitějším korporativním entitám.

V případě geografických entit - dat o sídlech Moravy a Slezska³ bylo zpracováno rozdělení na entity dílo/výtvor a geografické entity, ale dodatečně se ukázalo, že bude potřeba revidovat a sjednotit stručné charakteristiky i některé bližší určení zejména zaniklých obcí. Proto byl import dat pozastaven a probíhá (většinou už proběhla) úprava záznamů. V případě dat z území Čech⁴ se pracuje na konsolidaci dat, kteréžto práce se protáhnou do začátku roku 2016. Pracovalo se na propojení s databází CZ_RETRO.

V případě korporativních entit pokračovaly práce na popisu organizací z oblasti dopravy a některých korporací z oblasti okresní politické správy (byly již vyhotoveny prostřednictvím webového rozhraní INTERPI), na zahraničních služebních cestách byly zkoumány cizojazyčné názvy korporací (zejména z oblasti Uher a Slovenska). K vydání byla připravena monografie *Šefraná, Věra Adina, Přehled vybraných filmových korporací v českých zemích 1945-1970* vzniklá na základě dříve připravených a doplněných dat (viz příloha - výsledky). Publikace je v tisku.

V případě entit dílo/výtvor bylo pokračováno v doplňování údajů o vývoji názvů železničních stanic a zastávek a železničních tratí s cílem výsledky v rámci projektu publikovat. Práce se však velmi zpozdily nutností opakovaných revizí údajů, které se v úředních pramenech (věstníky, jízdní řády, služební pomůcky) vzájemně značně liší. Z těchto důvodů bylo prozatím vydání tiskem odloženo a bude realizováno už po ukončení projektu.

Naopak v případě entit typu osoba byla realizována publikace *Stupková, Marie - Klečacký, Martin, Slovník představitelů soudní správy v Čechách v letech 1849-1918, Praha : 2015, Masarykův ústav a Archiv AV ČR a Národní archiv, 464 s. ISBN 9788087782408* (srv. minulou průběžnou správu). Tato publikace však byla primárně vydána jiným vydavatelem a autory nejsou domácí tvůrci v duchu evidence RIV, proto ve výsledcích projektu INTERPI není uváděna. Projekt INTERPI je v publikaci výrazně zmíněn.

³ Hosák, Ladislav - Šrámek, Rudolf: Místní jména na Moravě a ve Slezsku I, II. Academia Praha, 1970, 1980

⁴ Profous, Antonín: Místní jména v Čechách. Jejich vznik, původní význam a změny. Nakladatelství Československé akademie věd, 1954-1960, díl I-IV

4 Příprava metodických a metodologických postupů, jejich zavedení v archivech

Již méně intenzivně než v minulém roce byla zejména začátkem roku prezentována Základní pravidla pro zpracování archiválií (dále též „Základní pravidla“) vydaná v roce 2014 Ministerstvem vnitra a schválená jako závazná pro použití v českých archivech⁵ na „roadshow“ jednotlivým veřejným archivům čtyřčlennou skupinou, z níž tři osoby jsou zapojeny do projektu INTERPI. Samostatná část prezentací byla věnována popisu původců archiválií a tvorbě přístupových bodů s ohledem na projekt INTERPI.

Výraznější pozornost byla věnována osvětě v rámci pracovní skupiny již zmíněného projektu Technologické agentury ČR „Vývoj software pro popis archiválií na základě Základních pravidel pro zpracování archiválií metodou PCP“, kde je napojení na INTERPI klíčovou funkcionalitou.

⁵ Metodický návod č. 1/2013 odboru archivní správy a spisové služby Ministerstva vnitra, kterým se vydávají nová Základní pravidla pro zpracování archiválií, č. j. MV-46913/AS-1/AS-2013 ze dne 17. dubna 2013