



***Příspěvek představuje projekt INTERPI – Interoperabilita v paměťových institucích, jehož cílem je vytvořit infrastrukturu pro tvorbu znalostní databáze o informačních objektech zpracovávaných v různých typech paměťových institucí – knihovnách, archivech, muzeích a galeriích. Přináší informace o dosavadních nejdůležitějších výsledcích projektu, mezi které patří konceptuální model a metodiky.***

## Úvod

Již několik let vnímají paměťové instituce (knihovny, archivy, muzea i galerie) změny v potřebách uživatelů a v jejich informačním chování, které vyplývají z měnícího se informačního prostředí. Vznikají iniciativy pro digitalizaci sbírek, pro vytvoření kvalitnějších metadat o sbírkových předmětech i pro jejich zpřístupnění ve virtuálním prostředí takovým způsobem, který by podporoval přirozené chování uživatelů a tvůrčí využívání zpřístupněných informací.

Navzdory tomu, že různé typy paměťových institucí jsou stejné svým posláním (shromažďovat, zpracovávat, uchovávat a zpřístupňovat objekty kulturního dědictví), rozdílné jsou v jeho naplňování, a to především kvůli různé úrovni využívání informačních technologií a standardizaci procesů a postupů. V době, kdy uživatelé nerozlišují, z jakého zdroje informace pocházejí, ale kladou důraz na to, aby byly komplexní a kvalitní, není žádoucí, aby knihovny, archivy, muzea, galerie nebo instituce památkové péče prezentovaly informace o kulturních objektech odděleně bez propojení a souvislostí.

Potřeba najít styčné body v informacích o kulturních objektech zpracovávaných a zpřístupňovaných v paměťových institucích a umožnit jejich efektivní tvorbu a využívání se odrazila v cílech a náplni projektu INTERPI.



## Vizitka projektu

Projekt Interoperabilita v paměťových institucích (INTERPI) je financován Ministerstvem kultury ČR v rámci Programu aplikovaného výzkumu a vývoje národní a kulturní identity (NAKI) pod identifikačním číslem DF1P01OVV023. Řešitelem – koordinátorem je Národní knihovna ČR a spoluřešitelem Národní archiv, což znamená, že i z organizačního hlediska je projekt postaven na spolupráci institucí v působnosti dvou ministerstev (kultury a vnitra). Projekt byl zahájen v roce 2011 a bude ukončen 31. 12. 2015.

## Cíl projektu a očekávané výsledky

V návrhu projektu byl jeho cíl stanoven následovně: „připravit vědeckou a technologickou infrastrukturu pro podporu zpracovávání, sdílení a využívání kulturního obsahu ve formě metadatových informací o informačních objektech zpřístupňovaných v paměťových institucích“. Hlavním očekávaným výsledkem projektu je poloprovaz infrastruktury pro budování znalostního modelu kulturního dědictví. Samotná infrastruktura, jejímž jádrem je znalostní databáze, zahrnuje software pro tvorbu a zpřístupňování dat o informačních objektech (resp. entitách), metodickou podporu pro tvorbu znalostního modelu a technickou podporu v podobě datového modelu, webových služeb atd.

## Souvislosti

V našem prostředí mají v oblasti standardizace a společného zpřístupňování fondů největší zkušenosti knihovny – především v podobě kooperativní tvorby národních autorit a využívání standardů jako AACR2 a MARC formáty. Zároveň byl v letech 2007 – 2011 Metodickým centrem pro informační technologie v muzejnictví při Moravském zemském muzeu v Brně (CITeM) realizován projekt „Národní autority v prostředí muzeí a galerií – interoperabilita s NK ČR“, který poukázal na potenciál národních autorit jako propojovacího prvku i pro muzejní sbírky a taktéž zdůraznil rozdílné potřeby muzeí a galerií, co se týče rozsahu a struktury informací uchovávaných o fyzických osobách.

Prakticky souběžně s realizací projektu INTERPI se stále větší důraz klade na koncept sémantického webu a zpřístupňování informací formou propojených dat (linked data). Tyto tendence podpořily myšlenku projektu a zároveň přinesly další požadavky na technologie, které v projektu budou využity.

## Konceptuální model – první a klíčový výsledek

Hned v úvodu řešení projektu bylo nutné sesbírat a analyzovat požadavky jednotlivých typů paměťových institucí na rozsah a strukturu informací o společných autoritách (?), objektech (?), entitách (!). I když projekt navazuje na myšlenku souborů autorit, bylo nevyhnutelné terminologii, tak blízkou knihovnickům, ihned opustit. Termín „autorita“ je pro ostatní paměťové instituce zatížen jinými významy a v knihovnickém chápání je jim cizí. Unifikované selekční prvky bez dalšího, dosti širokého, informačního aparátu a tvořené pouze ze zpracovávaných dokumentů jejich potřebám nevyhovují – očekávají větší informační hodnotu a víc souvislostí.

Klíčovým výsledkem projektu je tedy konceptuální model problematiky – tj. třídy a podtřídy entit, které vznikly na základě společných vlastností entit.

V projektu INTERPI tak rozlišujeme 7 tříd entit:

- osoba/bytost,
- rod/rodina,
- korporace,
- geografický objekt,
- událost,
- dílo/výtvor,
- obecný pojem.

Jako jeden z důležitých závěrů diskusí při tvorbě konceptuálního modelu je potřebné zmínit, že vytvoření systému tříd vyplývá z přirozené potřeby vymezit základní charakteristiky jednotlivých entit (tj. jaké informace a jakým způsobem se o entitě zaznamenávají); přitom systém tříd nesmí být pro popis entit limitou. Entita zařazená do jedné třídy může mít podle potřeb vlastnosti entit jiné třídy. V praxi to znamená například prolínání se tříd geografický objekt a dílo/výtvor pokud zpracováváme kulturní stavební památky, vodní díla apod.

## Metodika tvorby znalostního modelu INTERPI

Převedení konceptuálního modelu do konkrétní roviny pravidel a doporučení pro zaznamenávání vlastností entit znamenalo vytvoření Metodiky tvorby znalostního modelu INTERPI, která byla v závěru roku 2014 předložena k certifikaci Ministerstvu kultury.

Metodika obsahuje charakteristiku jednotlivých tříd entit a doporučení jak postupovat při výběru a zápisu vlastností entit. Vychází z pravidel využívaných v jednotlivých paměťových institucích (AACR2, RDA, CCO, Základní pravidla pro zpracování archiválií a další). Jejím cílem není stávající pravidla nahradit, ale na místech, kde se rozcházejí, navrhnout kompromis, nebo umožnit paralelní zápis vlastností v různých podobách. Typickým příkladem je koexistence více preferovaných forem jména, pokud je nutné vytvořit jeho jinou podobu s ohledem na různá pravidla. Metodika se zároveň snaží doplnit doporučení i pro ty třídy entit, které jsou v současných pravidlech rozpracovány nedostatečně nebo vůbec – například třída, dílo/výtvar nebo obecný pojem.

#### **Třída entit osoba/bytost**

Třída entit osoba/bytost zahrnuje skutečné a fiktivní fyzické osoby, bytosti s vlastním jménem evokujícím jméno osoby (např. božstva), zvířata známá pod vlastním jménem.

Ve třídě osoba/bytost bylo potřebné vyřešit, co představuje jedna entita v této třídě. Je to osoba v komplexním pojmání se všemi variantními jmény a identitami, nebo je to osoba ve smyslu bibliografické entity (tj. spojení různých forem jmen vyskytujících se v dokumentech se stejným tématem)? V metodice nabízíme dva způsoby jak entitu s více identitami zpracovat. První (a upřednostňovaný) postup spočívá ve zpracování všech identit do jednoho záznamu, přičemž jedna z forem jména se vybere jako preferovaná. Tento způsob je vhodný, jsou-li identity více méně stejné – opírají se o stejná biografická data. Druhý způsob umožňuje reálnou a fiktivní identitu zpracovat odděleně a navzájem je propojit, což je vhodné především v případě, že se v různých identitách využívají různá biografická data a panuje přítom nejistota, že jde o tutéž osobu. Další otázky vyplývají z výskytu kolektivních pseudonymů, anonymních a neznámých autorů. Tvůrci s kolektivním pseudonymem se zařazují do této třídy jako fiktivní osoby, pokud má jejich označení charakter jména osoby. Pro anonymní a neznámé tvůrce se využívají specifické formy označení a zařazují se rovněž do třídy entit osoba/bytost.

#### **Třída entit rod/rodina**

Tato třída zahrnuje skutečné a fiktivní rodiny, rody a jejich větve. V INTERPI jsme třídu vyčlenili, aby bylo umožněné komplexní zpracování genealogie.

#### **Třída entit korporace**

Korporace byly jednou z nejdiskutovanějších tříd z hlediska rozdílů v metodách jejich popisu v knihovnách a archivech. Například pravidla využívaná knihovnami nepovažují zkratky označující korporativnost (a. s., n. p., s. r. o. a jiné) za významné, kdežto v archivech jde o důležitou informaci. Metodika tvorby znalostního modelu proto přijala pravidla archivů a doporučuje rozlišovat jednotlivé korporace i podle typu.

Další závěry bylo nutné přijmout v souvislosti s vnímáním rozdílů mezi korporacemi, díly (stavby – např. hrady, zámky) a geografickými objekty (např. města). Je potřebné rozlišovat entity jako správní celky (spadající mezi korporace), pojmenoványými částmi přírody vytvořenými lidmi (spadající pod geografické objekty – např. parky) a stavbami (spadajícími pod entity dílo/výtvar).

#### **Třída entit geografický objekt**

Geografický objekt definujeme zjednodušeně jako entitu vymezenou stálým souborem geografických souřadnic. Vývoj geografického objektu je obvykle popsán v jednom záznamu. Specifikem znalostního modelu INTERPI je, že umožňuje popsat různé významné vývojové etapy geografického objektu odděleně a záznamy navzájem propojit příslušnými vazbami (např. historické správní obvody – okresy, kraje apod.).

#### **Třída entit událost**

Třída zahrnuje pojmenované dočasné události (dlouhodobé i krátkodobé, jednorázové i opakované) a taktéž pojmenované entity související s lidskou činností jako jsou zvyky lidové kultury, bitvy, konference a pojmenované dočasné přírodní jevy.

Metodika doporučuje každé opakování události popsat jako samostatnou entitu, přičemž lze současně vytvořit souborný záznam, který obsahuje informace vztahující se k historii celku. Jednotlivé události jsou tomuto soubornému záznamu podřazeny.

#### **Třída entit dílo/výtvar**

Jak již bylo zmíněno – entity dílo/výtvar představují skupinu, které byla doposud v odborných kruzích věnována malá pozornost. Koncept díla byl rozpracován v modelu FRBR, ale pro svou abstraktnost nebyl příliš aplikován. V prostředí muzeí a galerií se pod dílem chápe výsledek tvorby. Pro potřeby INTERPI jsme dílo vymezili na obecné úrovni jako výsledek záměrné lidské činnosti a jeho popis je zaměřen na zachycení aktuálního (posledního) stavu. Pokud je to nevyhnutelné, umožňuje znalostní model zachytit i jednotlivé významné historické etapy díla nebo jeho části a propojit tyto informace navzájem. Třída entit dílo/výtvar zahrnuje díla bez ohledu na to, zda v současnosti existuje nějaká jejich fyzická reprezentace.

### Třída entit obecný pojem

Zahrnuje označení kategorií osob a korporací, obecná označení věcí, zvířat, rostlin a abstraktní pojmy pro vlastnosti, oblasti poznání, umělecké styly apod. (např. oko, láska, mamut, dělník).

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

Paměťové instituce se přirozeně snaží různým způsobem unifikovat některé prvky popisu předmětů svého zájmu (sbírkových předmětů, knihovnických jednotek, archiválií – obecně kulturních objektů) tak, aby ulehčily samotný popis i zpřístupnění. Využívají pro tento účel více, či méně strukturované rejstříky, číselníky a řízené slovníky. Dalším výsledkem projektu jsou doporučení (v podobě metodiky) jak postupovat při mapování resp. harmonizaci zmíněných terminologických zdrojů tak, aby mohly být integrovány do znalostní databáze INTERPI. Metodika mapování a harmonizace terminologických zdrojů byla také předložena k certifikaci Ministerstvu kultury na konci roku 2014.

Zaměřuje se především na zdroje obsahující obecné pojmy – protože ty jsou v rámci paměťových institucí zpracovány na nejvyšší úrovni – jak obsahové (odborné), tak formální. Problémem jsou například rejstříky, které obsahují individuální i obecné pojmy (tj. jména osob, názvy institucí a geografických objektů spolu s termíny označujícími formu kulturního objektu), nebo řízené slovníky obsahující specifickou terminologii, ale bez rozlišení vícevýznamových termínů.

Ze specifických rysů terminologických zdrojů vyplývají následující problémy:

1. problémy s významem termínů – předpokládá se, že terminologické zdroje budou obsahovat:

- příliš specifické termíny, které se v databázi INTERPI nenacházejí,
- jednoslovné termíny obecného významu (operace),
- významově málo specifické termíny (život a dílo),
- příliš obecné, nerozlišené termíny používané namísto užších termínů – je potřeba poznamenat, že tento nedostatek lze odhalit pouze porovnáním na základě použití termínů;

2. problémy s formou termínů – terminologické zdroje mohou využívat při tvorbě termínů:

- jednotné/množné číslo,
- inverzi/přímé pořadí slov,
- velká/malá písmena,
- jinou pravopisnou normu,
- řetězce předmětových hesel (to znamená spojení více obecných hesel pro vyjádření specifického tématu).

Metodika obsahuje doporučení jak postupovat při předzpracování terminologických zdrojů a jaké metody harmonizace resp. mapování využít, aby se uvedené problémy minimalizovaly.

Základ obsahu (terminologický základ) znalostní báze INTERPI bude tvořit soubor národních autorit, přičemž v případě obecných pojmů bude pro tento účel doplněn o různé nepreferované podoby termínů – např. jednotné nebo množné číslo, invertovanou podobu termínu atd. Národní autority jsou jako terminologický základ vhodné především proto, že obsahují aktuální termíny, jsou tematicky univerzální, používají víceslovné deskriptory (tj. specifické a detailní termíny), obsahují notaci MDT (co usnadní specifikaci i provázání termínů) a obsahují anglické ekvivalenty (pro uplatnění výsledků na mezinárodní úrovni).

Pro ověření postupů doporučených Metodikou jsme experimentálně porovnávali termíny z číselníků systému DEMUS, klíčová slova ze systému LANius a Clavius a termíny z Polytematického strukturovaného hesláře se souborem věcných tematických autorit. Pro porovnávání jsme využili metriky Jaro a Jaro-Winkler. Výsledky poukazují na zmíněnou potřebu předzpracování termínů a doplnění terminologického základu databáze INTERPI.

Na obecné rovině podporuje znalostní model INTERPI mapování a harmonizaci terminologických zdrojů metodou korelace i metodou federace. Metoda korelace spočívá v přiřazení jednoho termínu k ekvivalentnímu termínu pocházejícímu z jiného zdroje. Pokud je to nezbytné, chybějící termín z jednoho zdroje se do druhého doplní. Tento postup je vhodný a možný pro mapování/harmonizaci zdrojů, které nejsou tematicky příliš vzdálené a jsou na stejné úrovni specifikace. Metoda federace znamená uvádění odborných termínů pod jedním obecnějším univerzálním termínem. V případě INTERPI se využití této metody předpokládá u vysoce specializovaných termínů (např. u regionálních variant označení pro stavby lidové architektury apod.), které se propojí s obecným termínem.

Metodika mapování a harmonizace terminologických zdrojů může být využita i v případě, že konkrétní paměťová instituce zatím žádný řízený slovník nevyužívá – obsahuje totiž doporučení pro výběr a reprezentaci pojmů.

## Software, transformace dat, příprava a úprava dat a další výsledky a činnosti

K zmíněným hlavním dosavadním výsledkům projektu je možno přičíst také vytvoření softwaru pro zpracování informací o entitách v souladu s Metodikou tvorby znalostního modelu. Software je realizován jako webové rozhraní a obsahuje formuláře pro různé třídy entit (resp. podtřídy entit). Největší výzvou se při tvorbě softwaru ukázalo zachování přehlednosti i při širokém spektru zpracovávaných informací a komplexnosti zaznamenávaných vztahů.

Dalšími činnostmi, které při řešení projektu nemůžeme opomenout, jsou přípravy a úpravy dat v Národní knihovně i Národním archivu. V Národní knihovně je to především doplnění geografických a tematických autorit, výběr a doplnění záznamů bitev a staveb, atd. V Národním archivu jsou zase připravovány příklady záznamů liniových staveb, zpracovávány záznamy korporací ústřední státní správy, doplňování geografických entit o historické názvy a jiné.

S vytvořením obsahového základu projektu INTERPI úzce souvisí také transformace dat z různých systémů – to znamená mapování stávajících využívaných formátů na datovou strukturu INTERPI, které je důležitým prvkem modelu kooperace.

## A co dál

V závěrečné etapě projektu INTERPI nás čeká především ověřování výsledků projektu a řešení technologických a organizačních otázek pro zabezpečení poloprovozu a hledání zanícených spolupracovníků pro jeho uvedení do života po skončení financování projektu.

Jedním z nejdůležitějších závěrů z průběhu projektu INTERPI je vůle paměťových institucí spolupracovat, hledat kompromisy, a také ustupovat ze svých požadavků, přičemž právě tento bod byl jedním z hlavních rizik projektu. V tomto směru lze již teď považovat projekt INTERPI za úspěšný.

**Ing. Nadežda Andrejčíková, PhD.**

andrejcikova@cosmotron.cz

(Cosmotron Slovakia, s. r. o.)

**Mgr. Marie Balíková**

marie.balikova@nkp.cz

(Národní knihovna České republiky)

**Ing. Miroslav Kunt**

kunt@nacr.cz

(Národní archiv)

**Mgr. Jana Šubová**

subova@cosmotron.cz

(Cosmotron Bohemia, s. r. o.)

