

## Možnosti a formy zpřístupnění archivních fondů nebo jejich součástí veřejnosti v elektronické podobě

Konference Archivy, knihovny a muzea v digitálním světě, 3.- 4. 12. 2008, Praha.

Miroslav Kunt / Národní archiv / e-mail: [miroslav.kunt@nacr.cz](mailto:miroslav.kunt@nacr.cz)

Ministerstvo vnitra v rámci svých projektů výzkumu a vývoje vypisuje i témata týkající se archivnictví. Jedním z nich je také tříletý projekt VE 20072009 zabývající se popisem archiválií a navazující problematikou digitalizace řešený Národním archivem ve spolupráci s dalšími subjekty. Hlavním cílem projektu je návrh implementace mezinárodních standardů (zejm. ISAD(G), ISAAR(CPF), EAD, EAC a METS) v českém prostředí tak, aby byly postupně přijaty jako základ práce archivů v ČR. S tím souvisí otázka pokusu o vytvoření jednotné platformy pro výměnu dat a také vazby na standardy používané ostatními paměťovými institucemi – muzei a knihovnami. Jedním z výstupů bude i jednoduché softwarové řešení prezentace archivních metadat (popisů archiválií) veřejnosti dálkovým přístupem a metodika pro digitalizaci archiválií. Důraz je tedy kladen na zpřístupnění archivních pomůcek a archiválií veřejnosti včetně provázání informací s knihovnami a muzei.

Aby měl projekt naději na úspěch při nasazení do stávající praxe, podílejí se na práci nejen odborníci z archivů, ale také z obou (konkurenčních) firem dodávajících archivní software. Dílčí výstupy jsou navíc konzultovány s odborníky z Odboru archivní správy a spisové služby MV, která je vlastně odborným garantem zadavatele projektu.

Strategickým cílem projektu je zastavit postupné odborné zaostávání českého archivnictví v oblasti zpřístupňování a umožnit jeho začleňování do mezinárodních projektů. Zásadní roli zde hraje zvyšující se míra standardizace v archivnictví ve světovém měřítku. Zmíněné zaostávání činí v současnosti přibližně pět let.

Nejtěžší práce – kromě dokončení rozpracovaných témat – proběhne v roce 2009: navrhnout nová pravidla pro zpracování archiválií s respektováním mezinárodních standardů. Ta stávající, pocházející z roku 1958, jsou archivní obcí považována za „svatá“ a všechny pokusy (od roku 1970) je změnit skončily nezdarem. Hraje zde přirozeně roli také psychologické hledisko a samostatnost archivů. Jestliže naše knihovny jsou schopny akceptovat jednotné standardy a považují dnes elektronickou formu popisu za samozřejmou, v českých archivech tomu tak zdaleka není. To se ovšem zakrátko změní. Navržená novela vyhlášky 645/2004 Sb., kterou se provádí některá ustanovení zákona o archivnictví a spisové službě a o změně některých zákonů výslovně uvádí, že kromě listinné formy je nutné archivní inventář (základní archivní pomůcka) zasílat do celostátní evidence i v elektronické podobě, a to v XML schématu stanoveném Ministerstvem vnitra.

### Mezinárodní archivní popisné standardy

Mezinárodní archivní rada (<http://www.ica.org>) přijala v roce 1994 (1. verze), resp. 1999 (2. verze) „Všeobecný mezinárodní standard archivního popisu ISAD(G)“ a následně též „Mezinárodní standard záznamů archivních autorit pro korporace, osoby a rody ISAAR(CPF)“, jehož 2. vydání bylo připraveno v roce 2004. Tyto standardy dosud neměly oficiální český překlad (u ISAAR(CPF) neexistoval překlad žádný). V rámci projektu byly tedy provedeny odborné překlady obou standardů (posledních verzí) včetně vytvoření příkladů popisu. Ještě v březnu budou oba překlady předány Ministerstvu vnitra jako členskému subjektu k oficiálnímu zaslání Mezinárodní archivní radě. Komentované zveřejnění se předpokládá v průběhu roku 2009.

Na popisných standardech je cenné především to, že pojmenovávají některé „zcela samozřejmé“ pojmy a postupy, které ovšem nejsou v práci archivů dostatečně akcentovány.

Dle ISAD(G) „Cílem archivního popisu je určit a vysvětlit obsah archivního materiálu včetně všech souvislostí tak, aby se usnadnila jeho dostupnost (přístupnost). Dosáhnout tohoto cíle lze vytvářením přesných a odpovídajících vyjádření (popisů) a jejich uspořádáním podle předem daných modelů.“ Základem těchto „modelů“ je vždy víceúrovňový (hierarchický) popis, který se částečně podobá analytickému popisu v knihovnách. Jeho základními postupy jsou:

- 1) Popis od obecného ke specifickému – znázornit souvislosti (kontext) a hierarchickou strukturu fondu a jeho částí.
- 2) Informace odpovídající úrovni popisu – znázornit odpovídající souvislosti (kontext) a obsah jednotky popisu.
- 3) Propojení popisů – určit jednoznačné umístění (pozici) jednotky popisu v hierarchii.
- 4) Neopakování informací – vyhnout se nadbytečnosti (redundanci) informací v hierarchicky uspořádaném archivním popisu.

Standard ISAD(G) je standardem obecným, tzn. nestanovuje pravidla pro popis jednotlivých dokumentů nebo jejich druhů, ale určuje strukturu archivních pomůcek. Jeho pravidla jsou uspořádána do sedmi oblastí:

- 1) Identifikační údaje – základní informace k identifikaci jednotky popisu.
- 2) Souvislosti (kontext, vztahy) – informace o původci a úschově jednotky popisu.
- 3) Obsah a struktura – informace o podstatě a uspořádání jednotky popisu.
- 4) Podmínky přístupnosti a využití – informace o dostupnosti jednotky popisu.
- 5) Doplňující materiál – informace o dalším materiálu, který významně souvisí s jednotkou popisu.
- 6) Poznámky – informace speciální a takové, které nemohou být uvedeny v žádné z ostatních oblastí.
- 7) Kontrola popisu – informace o tom, jak, kdy a kým byl archivní popis vytvořen.

Standard ISAAR(CPF) byl nepochybně inspirován tvorbou autoritních záznamů v knihovnách, ale jeho cíle jsou v zásadě jiné, což je zřetelné hned v úvodní kapitole:

„Popis původce je základní činností archiváře, bez ohledu na to, zda jsou popisy spravovány manuálně nebo v automatizovaných systémech. To vyžaduje plnou dokumentaci a trvalé udržování údajů o kontextu vzniku a používání dokumentů, z nichž nejdůležitější je provenience. Archivní autoritní záznamy mohou být použity:

- a) k popisu korporace, osoby nebo rodu jako jednotek v systému archivního popisu a/ nebo
- b) k řízení vytváření a využívání přístupových bodů v archivním popisu,
- c) k dokumentování vztahů mezi různými původci dokumentů, vztahů mezi nimi a dokumenty či jinými zdroji jimi vytvořenými a/nebo vytvořenými o nich.“

Je zřejmé, že s knihovnickou definicí autoritního záznamu archiváři nevystačí. Ostatně nejen archiváři. Například návrh nových datových polí pro autoritní záznamy v MARC21, který zveřejnila Kongresová knihovna 9. ledna 2009 (<http://www.loc.gov/marc/marbi/2009/2009-01-1.html>), zcela jasně vychází ze standardu ISAAR(CPF) a archivních zvyklostí popisu původců.

Obdobně jako ISAD(G), jsou v ISAAR(CPF) informace rozčleněny do čtyřech oblastí:

- 1) Oblast identity – zde je informace podávána tak, aby jedinečným způsobem identifikovala záznamem popisovanou entitu a definovala standardizované přístupové body používané v dokumentech).
- 2) Oblast popisu – příslušná informace týkající se podstaty, kontextu a aktivit popisované entity.
- 3) Oblast vztahů – zaznamenání a popis vztahů k jiným korporacím, osobám a/nebo rodům.
- 4) Oblast kontroly – jedinečná identifikace autoritního záznamu a zaznamenání informace o tom, kdy a kým byl záznam vytvořen a spravován.

V Českém archivnictví dosud neexistuje vhodný popis původců. Je součástí první kapitoly tzv. úvodu k archivní pomůcce, ale jednotliví původci zde nejsou separováni a text je často orientován na popisované archiválie, nikoli na původce samé.

Autoritní záznamy o původcích mají velký význam pro kontextuální nadstavbu archivního popisu – lze jimi vytvořit interaktivní „dějiny správy“ jako další přístupový bod k archiváliím, původce lze vázat nejen na celé fondy, ale i na jejich části (např. u prior1, sérií apod.). Nejlépe je tato problematika zvládnutá v australských archivech (srv. Public Record Office Victoria (PROV) <http://www.access.prov.vic.gov.au> nebo State Records Authority of New South Wales <http://investigator.records.nsw.gov.au>).

Obdobný přístup k autoritním záznamům jako archivy uplatňují muzea a galerie, kde je oblast řešena výzkumným projektem DC07P02OUK002 brněnského Metodického centra pro informační technologie v muzejnictví při Moravském zemském muzeu v Brně (CITEM) „Národní autority v prostředí muzeí a galerií – interoperabilita s NK ČR“. Analýza a možnost využití databáze Národních autorit je ostatně i jedním z cílů zde popisovaného projektu.

Dnes již existují další dva mezinárodní archivní popisné standardy:

- ISDIAH (International Standard for Describing Institutions with Archival Holding) pro popis pamětových institucí spravujících (nejenom) archiválie,
- ISDF (International Standard for Describing Functions) pro kontextové údaje vyjadřující působnost původce ve vztahu v jím vytvořeném archiváliím. Funkce znamenají nejčastěji do podoby tezauru uspořádané agendy svěřené legislativou, normami apod. a obecné pojmy charakterizující činnost původce (srv. <http://www.access.prov.vic.gov.au>, např. „státní uhelné doly“, „dolování“ nebo „železnice“).

Tyto dva novější standardy nejsou součástí projektu, ale použití ISDIAH se předpokládá. ISDF je v našich podmínkách organizační nestability původců jen velmi obtížně uplatnitelný – alespoň do doby, než bude vytvořena funkční vrstva archivních autoritních záznamů. Skutečně vzorové použití funkcí jako přístupových bodů ukazují výše citované australské archivy.

## Mezinárodní archivní technické standardy

Podobně jako u knihoven v případě formátů MARC, jsou pro archivy mezinárodně standardizovány technické datové formáty vycházející z popisných standardů a vytvořené prostřednictvím jazyka XML.

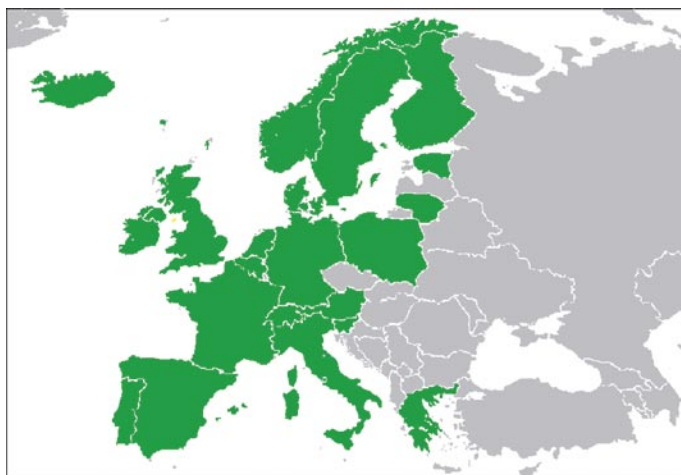
Z popisného standardu ISAD(G) vychází EAD (Encoded Archival Description) a z IS-AAR(CPF) EAC (Encoded Archival Context). Zatímco EAD, udržovaný podobně jako MARC Kongresovou knihovnou USA (<http://www.loc.gov/ead/ead.html>), dosáhl již velkého rozšíření, v případě EAC je zveřejněna pouze testovací verze (<http://www.iath.virginia.edu/eac/>) a je nutné přiznat, že tento standard má určité mezery. Datový XML standard by například neměl definovat a používat formátovací tagy (např. `<p></p>` pro definici odstavce).

V České republice vznikl v letech 2005-2006 národní XML „Standard pro ukládání a zasílání archivních pomůcek druhu inventář a dílčí inventář v digitální podobě“ a „Standard pro jednotnou evidenci matrik a výměnný formát pro ukládání a zasílání záznamů matrik v digitální podobě“ (dostupné z <http://www.mvcr.cz>). V současné době je potřeba si přiznat, že není možné zejména s prvně jmenovaným standardem vystačit a bude nutné převzít standard mezinárodní – EAD. Obdobně také naše národní rozlišování různých druhů archivních pomůcek (podrobnosti v příloze 2 vyhlášky 645/2004 Sb.) by nemělo mít vliv na použití standardů a využití pomůcek uživateli.

Na americkém kontinentě (hlavně v USA a Kanadě) se pro popis archivních fondů a sbírek používal v minulosti speciální MARC AMC, který vznikl v roce 1984. MARC AMC neumožňoval popis jednotlivých částí archivního materiálu tak, jak jsme zvyklí v našich podmínkách, ale jen souhrnný popis celého archivního souboru v rámci knihovního katalogu. Používal se do konce 90. let, kdy ho nahradil EAD, ale vazba na knihovní katalogy přes formát MARC dále zůstala. Je vyjádřena možností mapování na pole MARC nebo jiných standardů prostřednictvím atributu „encodinganalog“:

```
<corpname encodinganalog="MARC 110">Národní archiv</corpname>.
```

Knihovny v anglosaských zemích mají často jiný charakter než u nás, jsou současně i archivy (příkladem je Kongresová knihovna ve Washingtonu). Zmíněné mapování je ve standardech zcela nativně, evidentně z historických důvodů pro dodržení kompatibility. U nás můžeme využít tuto užitečnou vlastnost pro „porozumění“ mezi systémy archivními a knihovními.



Již dnes jsou standardy EAD a EAC požadovány jako podmínka spolupráce např. v rozvíjejícím se evropském projektu APEnet: Archival Portal of Europe (<http://www.european-archivesgateway.eu/>); ověřené rozšíření EAD v Evropě je zřejmé z obrázku:

## Jiné mezinárodní standardy

V archivech se kromě úředních knih, spisového a účetního materiálu nacházejí i jiné typy dokumentů archivní povahy (např. rukopisy, prvotisky, staré tisky, noviny, výroční zprávy, mapy a plány, grafické listy, fotografie, hudebniny aj.), které zároveň spravují i jiné paměťové instituce (knihovny, muzea a galerie), používající odlišné způsoby popisu těchto dokumentů. Zatímco se knihovníci přidržují pravidel AACR2 (Anglo-American Cataloguing Rules, Second Edition), která jsou založena na popisu jednotlivých dokumentů bez vyjádření vazby k celému souboru a celý tento popis je reprezentován ve výměnném formátu MARC (Machine-Readable Cataloging), muzejníci vycházejí z ontologického modelu CIDOC CRM (Comité International pour la Documentation Conceptual Reference Model). Jde opět o znázornění formy, které umožňuje vytvářet a identifikovat vztahy mezi věcnými skupinami popisu a lépe tak definovat význam zapsaných hodnot (obsahu) popisu. Aby bylo možné vzájemně sdílet a vyměňovat si popisy mezi jednotlivými institucemi, vznikla otevřená iniciativa Dublin Core (DC). Ta se soustředila pouze na elementární položky popisu jakéhokoli informačního zdroje, které sjednotila a standardizovala. Protože toto řešení stále není ideální, vznikl další projekt, který je kompromisem mezi komplexním formátem MARC a jednoduchým Dublin Core, tzv. MODS (Metadata Object Description Standard). Specifické formáty, např. MASTER (Manuscript Access through Standards for Electronic Records) pro podrobný popis středověkých rukopisů a CEI (Charter Encoding Initiative) pro popis středověkých listin, neupřednostňují žádné oborové hledisko a jsou tak společnými formáty pro jakoukoli paměťovou instituci uchovávající daný typ dokumentů. Názvosloví jejich popisných položek vychází stejně jako v případě archivního standardu EAD z iniciativy TEI (Text Encoding Initiative). Analýza provedená v rámci projektu se snaží zhodnotit přínos jednotlivých standardů pro popis archiválií, zejména v oblasti identifikace dokumentů, vyjádření kontextu a ve způsobu znázornění formy popisu. Výsledkem práce bude mimo jiné i stanovení konverzních tabulek některých standardů, jejichž definice zajistí oboustranný převod.

K dalším technickým standardům, kterými se projekt zabývá, patří standardy kontejnerové a komunikační.

Posuzován byl kontejnerový standard METS (Metadata Encoding and Transmission Standard, <http://www.loc.gov/standards/mets/>), který umožňuje zapouzdření digitálních objektů spolu s jejich metadaty za účelem jejich přenosu. V digitálních archivech (i v projektu našeho Národního digitálního archivu) je využíván pro tvorbu tzv. informačních balíčků jako základního prvku digitální archivace. V popisovaném projektu byl sledán jako klíčový i pro přenos a správu digitálních reprodukcí a jejich metadat (viz dále).

Samostatnou kapitolou jsou komunikační protokoly. Pro účely sklizení (harvestování) metadat ve formě celých archivních pomůcek se předpokládá využití OAI-PMH, pro komunikaci pak použití webových služeb SRU/SRW. Knihovnický protokol Z39.50 je pro účely archivnictví nedostatečný, předpokládá využití kódování MARC a jeho implementace je obtížná a drahá. Jeho využití projekt předpokládá pouze pro zajištění napojení na jiné paměťové instituce, např. mezi databázemi národních autorit spravovanou Národní knihovnou ČR a vrstvou archivních autorit.

## Problematika digitalizace

V současnosti používané programy pro popis archiválií váží digitalizované dokumenty k popisu různými způsoby, např. prostřednictvím identifikátoru odvozeného z čísla archivu a archivního fondu. Není uspokojivě vyřešeno připojení a prezentace datově rozsáhlých objektů a objektů jiného charakteru než rastrová grafika. Vlastní digitalizované soubory

jsou odděleny od popisu a většinou k němu připojeny velmi volně. Zachovat vazby mezi popisem a digitalizovanými objekty je v archivu značně obtížnější než v knihovnách. Jednotkou popisu zde ve většině případů není dokument, ale skupina dokumentů, někdy značně rozsáhlá. Právě tyto problémy se podařilo minimalizovat při hromadné digitalizaci financované z Norských finančních mechanismů v Archivu hlavního města Prahy (<http://www.ahmp.cz>). Takto vyzkoušená metodika byla pro potřeby zde popisovaného projektu upravena a transformována. Z důvodů dlouhodobého uchování elektronických dokumentů (a bezpečnosti vazeb mezi digitalizovanou archiválií a popisem) se předpokládá dále využití již zmíněného kontejneru METS. V případě objemných datových souborů rastrové grafiky předpokládáme pro prezentaci využití technologie na bázi pyramidového TIFF, resp. JPEG s použitím open source software, např. Zoomify ([www.zoomify.com](http://www.zoomify.com)). Obraz je zde rozložen na čtverce pro každé zvolené zvětšení. Vrstvy jednotlivých zvětšení dohromady tvoří „pyramidu“: v „základech“ je největší velikost snímku, na „špičce“ nejmenší.

## Prezentace archivních pomůcek veřejnosti

V současné době jsou elektronické archivní pomůcky i data celostátní evidence archivních fondů (nejvyšší úroveň popisu) zaslány do druhotné a centrální evidence prostřednictvím e-mailů, disket apod., a to vždy jednou ročně. Jednotlivé archivy samozřejmě data zveřejňují různým způsobem ve svých informačních systémech. Tyto postupy mají zásadní nevýhodu – vyhledávání v heterogenním prostředí různých informačních systémů při nemožnosti zadat jeden dotaz do všech těchto systémů.

Projekt nechce a nebude konkurovat sofistikovaným systémům umožňujícím vedle zobrazení popisu archiválií a jejich reprodukcí např. také jejich objednání do studovny. Již existující řešení umožňují mapování registrovaných databází, sklizení (většinou prostřednictvím protokolů Z39.50 nebo OAI-PMH) a zpracování popisných metadat z nich na jednom místě. Příkladem je knihovnická Jednotná informační brána (<http://www.jib.cz>). Nedostatek tohoto řešení spočívá v tom, že takto zpracovaná data jsou přístupná zase jen z příslušného portálu, nikoli všeobecně pro nestructurované vyhledávání univerzálními vyhledávači (Google apod.).

Cílem projektu je:

- 1) Navrhnout, popř. vyzkoušet, sklizení metadat archivních pomůcek a popisů archivních fondů na jedno místo prostřednictvím OAI-PMH a SRU/SRW; sklizeny (harvestovány) budou celé objekty na základě metadat reprezentujících základní údaje o objektu (nekvalifikovaný Dublin Core).
- 2) Ze získaných údajů na základě daného XML schématu vytvořit statické HTML stránky, které po napojení na kmenovou stránku (index) budou mapovány univerzálními prohlížeči.
- 3) Vhodně zvolenou adresářovou strukturou takto vytvořených stránek a použitím funkcí některého univerzálního vyhledávače umožnit strukturované vyhledávání v datech.

Popsaná metoda je velice primitivní, ale také velmi účinná, odstraňující problémy s nutnou specifikací sklizených dat a umožňující zapojení dalších objektů definovaných XML schématem (např. z jiných paměťových institucí). To nevyklučuje budoucí tvorbu jednotného informačního portálu na základě shromážděných dat.

Obdobné řešení zvolily ve svém projektu německé archivy (<http://www.daofind.de>).

## Závěr

Projekt se snaží posunout české archivnictví dále. Některé uvažované postupy jsou v rámci řešení ve světě zcela nové nebo teprve připravované (např. využití standardu METS v archivech Německa, Švýcarska a Austrálie). Zda bude projekt úspěšný závisí nejen na implementaci řešení v software, ale především na ochotě archivářů přijmout nové postupy. Autor děkuje kolegovi Mgr. Tomáši Dvořákovi za cenné připomínky a obrázek rozšíření standardu EAD.

1. priora = spisy o předchozím jednání o téže věci, vyhledané a připojené registraturou k podání