

Úvod: Historie digitalizace, dlouhodobé ochrany a zpřístupnění archivních dokumentů v ČR

Nárůst používání počítačů v běžném životě je patrný ve všech oborech lidské činnosti a nelze před ním zavírat oči. S používáním výpočetní techniky velmi úzce souvisí i potřeba archivace digitálních dokumentů. Obecně stále ještě panuje nedůvěra k ukládání, uchovávání a následnému zpřístupnění digitálních dat. Ostatně ruku na srdce, kdo z nás ještě nepřišel o svá data? Poslední dobou roste i tlak veřejné správy na možnost předávat svá data v digitální podobě, neboť tisknout velké soubory jen kvůli archivaci, se zdá již přežitkem. Vznikají a rozvíjí se projekty typu e-justice, e-government, v jejichž rámci vzniká množství digitálních dokumentů, které je třeba uchovat pro budoucnost. Je zřejmé, že nejen o papírové, ale i o digitální dokumenty je nutně se postarat. Vzhledem k tomu, že péče o papírové dokumenty spadá do činnosti archivářů, připadla péče o digitální dokumenty rovněž archivům. Bylo rozhodnuto řešit tuto problematiku, alespoň v začátcích, centrálně, a proto při Národním archivu v Praze vzniká Pracoviště pro dlouhodobé uchovávání a zpřístupňování dokumentů v digitální podobě nazývané též Národní digitální archiv (NDA). Jeho úkolem je mimo jiné najít odpověď na otázku: Jak nakládat s digitálními dokumenty, aby byly i s odstupem času dostupné, čitelné a důvěryhodné?

První úvahy o možnostech digitální archivace se v českém archivnictví objevily přibližně v polovině devadesátých let 20. století. Na ně navázaly konkrétní výzkumné projekty realizované za účasti Odboru archivní správy MV, ČVUT a Národního archivu v letech 2001-2005. Dílčí zprávy byly publikovány na stránkách Archivního časopisu.

Konkrétní podobu pak úvahy o digitální archivaci dostaly v usnesení vlády č. 11 ze 7. ledna 2004. Tímto usnesením vláda uložila místopředsedovi vlády a ministru vnitra zpracovat ve spolupráci s ministrem informatiky projekt dlouhodobého uchovávání a zpřístupňování dokumentů v digitální podobě. V předkládací zprávě usnesení byla nejen konstatována a zdůvodněna nutnost řešit dlouhodobé uchovávání digitálních dokumentů, ale rovněž byl určen postup směřující k vybudování digitálního archivu. Nejprve měl být při Národním archivu sestaven tým, který připraví podklady pro zpracování projektu. Na jeho základě pak měl být vybudován digitální archiv.

Naléhavost řešení dlouhodobého uchovávání je zdůrazněna rozvojem e-governmentu, který není bez vyřešení digitální archivace v zásadě možný. Existenci digitálního archivu předpokládá například Národní plán zavedení elektronického zadávání veřejných zakázek pro období let 2006 až 2010 (vládní usnesení č. 500 z 10. května 2006).

Výše zmíněný tým pro přípravu digitálního archivu byl v Národním archivu ustaven na konci roku 2005. V průběhu roku 2006 shromáždil podklady a vypracoval zadávací dokumentaci. Na jejím základě Národní archiv vyhlásil výběrové řízení na zpracovatele projektu digitálního archivu. Vítězem se stala firma ICZ, a.s., která následně (v období od července 2007 do února 2009) zpracovala technologický projekt Pracoviště pro dlouhodobé uchovávání a zpřístupňování dokumentů v digitální podobě. Projekt byl dokončen a předán v březnu 2008 a je k dispozici na webových stránkách Národního archivu (http://www.nacr.cz/zpravy/projekt_nda.aspx).

Národní digitální archiv jako strategický projektový záměr pro čerpání finančních prostředků ze strukturálních fondů EU v rámci Smart Administration

V dubnu 2008 vláda svým usnesením č. 447, k zabezpečení plnění úkolů ve věci vybudování Národního digitálního archivu, schválila čerpání finančních prostředků na vybudování Národního digitálního archivu a uložila ministru vnitra nárokovat finanční prostředky ve výši 85 % nákladů projektu z evropských strukturálních fondů v rámci Integrovaného operačního programu. Následně pak byl projekt Národního digitálního archivu uveden v příloze vládního usnesení č. 536 ze 14. května 2008, o strategických projektových záměrech pro čerpání finančních prostředků ze strukturálních fondů Evropské unie v rámci Smart Administration – příloha ke strategii Efektivní veřejná správa a přátelské veřejné služby.

Klíčové postavení Národního archivu v rámci koncepce dlouhodobého uchování digitálních dokumentů

Problematika dlouhodobého uchování digitálních dokumentů je finančně velmi náročná. Je nutno očekávat náklady na vybudování několika datových úložišť, na tvorbu migračních, eventuálně emulačních nástrojů, na definici ukládacích strategií, datová spojení a v neposlední řadě i na výzkumnou činnost. S rostoucím počtem uchovávaných digitálních dokumentů budou tyto náklady relativně klesat, nicméně zpočátku budou značné (v řádu stovek milionů Kč). Proto současná koncepce předpokládá, že digitální archiv bude zřízen při Národním archivu, bude mít celorepublikovou působnost a bude poskytovat akreditovaným archivům službu dlouhodobého uchování dokumentů v digitální podobě. Případná regionální pracoviště budou vznikat až na základě zkušeností s provozem centrálního digitálního archivu.

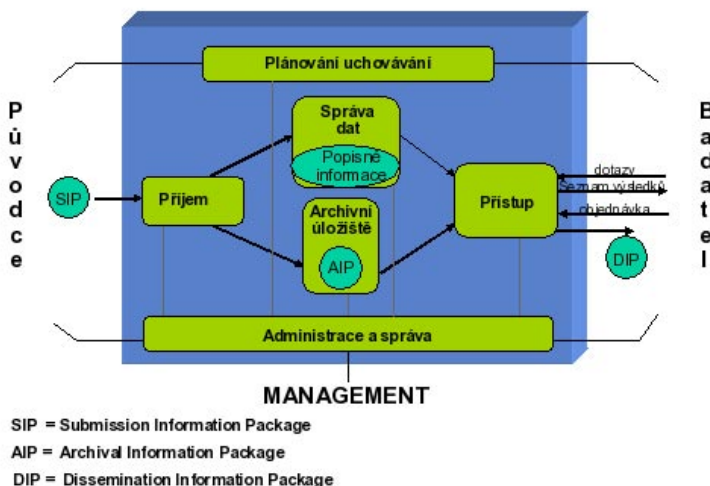
Primární funkcí digitálního archivu bude ukládat odborně vybrané dokumenty (archiválie), zajistit jejich dlouhodobé uložení (nad 50 let), jejich čitelnost a přístupnost při zachování integrity a nezbytné míry autenticity. Kromě uvedené primární funkce bude digitální archiv schopen uložit a zpřístupnit digitalizované archiválie (originály existují v čitelné analogové podobě) uložené v akreditovaných archivech.

Jak je zřejmé, je funkce Národního digitálního archivu omezena na dlouhodobé uchování a zpřístupňování dokumentů v digitální podobě a nepředpokládá se, že by digitální archiv sám prováděl výběr archiválií u původců. Odborné posouzení dokumentů, jejich případný výběr a archivní zpracování zůstává plně v kompetenci příslušných archivářů stejně, jako je tomu u tradičních archiválií. Předpokládá se, že všichni archiváři provádějící skartační řízení u původců budou pracovníky digitálního archivu dostatečně proškoleni, aby zvládli digitální dokumenty bez problémů přebírat. V případě komplikací jim budou pracovníci NDA plně k dispozici.

Hlavní funkce a architektura Národního digitálního archivu, předpokládané objemy archivovaných dat

Navržené řešení digitálního archivu vychází ze standardu OAIS (ISO 14721:2003 – Open Archival Information System). Tento standard vymezuje základní koncepci archivu pro uložení elektronických dokumentů a je na něm budována většina současných digitálních archivů (Velká Británie, Austrálie, USA). Standard definuje hlavní funkce, které má archiv zajišťovat. Jedná se o příjem, správu dat, archivní uložení, přístup, administraci a plánování uchování. Základní model OAIS je na obrázku 1.

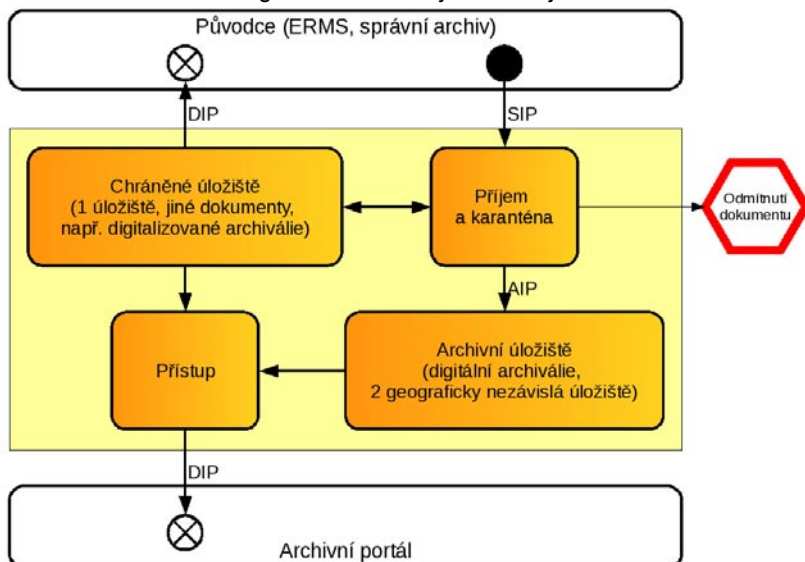
OAIS Funkční celky



Obr. č. 1

Podle standardu OAIS jsou elektronický dokument a všechny informace (metadata) zabaleny do balíčku s jednotnou strukturou. Podle standardu OAIS jsou tyto balíčky nazývány SIP – Submission Information Package (balíčky přijímané od původců), AIP – Archival Information Package (archivní balíčky zahrnující ukládaný obsah a jeho příslušné popisné informace pro uchovávání – metadata) a DIP – Dissemination Information Package (balíčky vytvořené na základě badatelského dotazu, pro využívání).

Základní navržené schéma digitálního archivu je následující:



Obr. č. 2

Digitální dokument, který bude archivářem vybrán za archiválii, bude nejprve připraven u původce do podoby vhodné pro předání do archivu, tzn. bude vytvořen balíček SIP. Balíček bude většinou automaticky generován elektronickým systémem spisové služby. Do balíčku bude zahrnut nejen dokument, ale i metadata, která se k němu váží a která byla pořizována během celého dosavadního životního cyklu dokumentu.

Zaslaný balíček bude v archivu nejprve umístěn do karantény. To znamená, že bude zkontrolován, zda soubory neobsahují škodlivý kód (viry apod.). Kontrola proběhne dvakrát s časovým odstupem třiceti dnů. Pokud dokument nebude obsahovat škodlivý kód, proběhne další kontrola balíčku. Zejména bude kontrolován formát souborů a rozsah vyplnění metadat. Jestliže balíček bude obsahovat škodlivý kód, nevhodný formát souborů či nebude obsahovat požadovaná metadata, bude odmítnut a původce bude o této skutečnosti informován.

Přijatý balíček bude dále zpracováván. Zejména budou doplněna metadata podporující procesy řízení uchovávání a zpřístupňování a dokumentu bude v rámci digitálního archivu přidělen jednoznačný identifikátor. Výsledkem celého procesu bude archivní informační balíček – AIP (všechna metadata a vlastní digitální soubory). Tento balíček bude následně uložen do archivního úložiště s řízeným přístupem, aplikací uchovávacích metod apod.

V průběhu prací na technologickém projektu bylo potřeba řešit otázku dokumentů, které nejsou digitálními archiváliemi v pravém slova smyslu, ale které by bylo vhodné rovněž uchovat. Jako řešení byl do schématu archivu zařazen meziklad digitálních dokumentů, takzvané Chráněné úložiště. Zejména v počátcích digitálního archivu zde mohou být uchovávány balíčky, které původce zaslal do archivu, ale které nesplňují předepsaná kritéria. Takový balíček bude původci buď rovnou vrácen nebo bude uložen v úložišti a původce ho ve spolupráci s archivem opraví (doplní). Též může jít o dokumenty – potenciální archiválie – s dlouhými skartačními lhůtami, které archiv pomůže původci dlouhodobě uchovat. Dalším způsobem využití úložiště je uchování digitálních obrazů klasických archiválií.

Je nutno zdůraznit, že do archivního úložiště ani do chráněného úložiště nebude možno ukládat nepopsané dokumenty. Každý dokument musí být opatřen alespoň základními metadaty (původce, identifikátor, datum vzniku, název, spisový plán, obsah, informace na základě čeho je ukládán). V počátcích digitálního archivu budou pravidla pravděpodobně benevolentnější, ale ani tehdy nebude možný prostý přesun dat z disku u původce do digitálního archivu.

Nedílnou součástí digitálního archivu je zpřístupnění uložených dokumentů. Zpřístupnění bude probíhat pomocí webového portálu Národního digitálního archivu, jehož prostřednictvím zadá uživatel svůj požadavek. Modul přístup následně žádost zpracuje, vytvoří DIP (Dissemination Information Package), který bude uživateli zobrazen prostřednictvím webového portálu. DIP může v závislosti na uživatelském požadavku obsahovat seznam přístupných dokumentů a jejich náhledy, konkrétní dokument a jeho metadata, informaci o eventuálním omezení přístupnosti dokumentu apod.

Způsob uchování

Na rozdíl od archivace klasických dokumentů, kdy stačí dokument uchovat a jsme schopni ho číst, je způsob uchování digitálních dokumentů komplikovanější. Pro interpretaci digitálního dokumentu vždy potřebujeme technický prostředek, který dokument převede z binární do čitelné podoby. Při uchování dokumentu tedy musíme zajistit nejen vlastní zachování dokumentu (dokument máme), ale i jeho čitelnost (dokument jsme schopni interpretovat).

Zachování dokumentu

Proti ztrátě bude dokument v archivním úložišti zajištěn uložením na dvou geograficky oddělených místech za použití dvou typů médií s rozdílným fyzikálním principem. Doku-

menty v chráněném úložišti pak budou uloženy na jednom místě a dvou médiích. Hlavní pracoviště obsahující chráněné úložiště a primární archivní úložiště digitálního archivu bude nově vybudováno v Praze v sousedství Národního archivu. Záložní pracoviště bude obsahovat sekundární archivní úložiště a pro jeho umístění bude využit stávající objekt. Obě úložiště budou od sebe vzdálena 100 km. Jako ukládací média archivních úložišť jsou v technologickém projektu navržena disková pole (magnetický princip záznamu) a UDO disky (optický princip záznamu) a jako ukládací média chráněného úložiště technologický projekt uvádí diskové pole zálohované na pásky.

Čitelnost dokumentu, formáty, migrace a emulace

Prosté zachování dokumentu ještě nezaručuje, že dokument bude možno interpretovat (číst). Vývoj formátů souborů používaných k záznamu informací je velmi dynamický, což přináší pro dlouhodobé uchování problémy. Z možných řešení zajištění dlouhodobé čitelnosti dokumentů uložených v digitálním archivu (emulace, migrace, virtualizace) byla zvolena metoda migrace. Jejím principem je přizpůsobení dat prostředí. Dokumenty uložené ve formátech, které nebudeme schopni interpretovat, se převedou do jiného (vhodnějšího) formátu. Nelze samozřejmě předpokládat, že v současné době lze touto metodou řešit dlouhodobé uchování veškerých digitálních dokumentů. To ostatně nedokáže zajistit žádná metoda. Migrace se však aktuálně jeví jako velmi efektivní způsob zachování čitelnosti většiny digitálních dokumentů. Prostřednictvím relativně levných migračních nástrojů lze dosáhnout převodu mnoha formátů primárních dokumentů do menšího množství formátů vhodných pro dlouhodobé uchování. Čím bude množství formátů menší, tím bude dlouhodobé uchování snazší, byť existuje určité riziko v případě, že formát pro dlouhodobé uchování bude nevhodně zvolen.

Pro každý formát dokumentu přebíraného do archivu musí mít digitální archiv vypracovanou uchovací strategii. Pokud taková strategie neexistuje neměl by být dokument převzat. V technologickém projektu jsou posouzeny některé formáty a došlo k jejich rozdělení do tří skupin: preferované, akceptované, neakceptované (s nízkou životností). Při posuzování formátů byla zohledněna tato kritéria: otevřenost, rozšířenost, transparentnost (jednoduchost uložení informací v dokumentu), sebedokumentace, vnější závislosti, vliv patentů, ochranné mechanismy. Rozdělení do jednotlivých skupin je patrné z tabulky č. 1.

typ dokumentu	preferované formáty	akceptovatelné formáty	formáty s nízkou trvanlivostí
textový dokument	prostý text, XML struktura, PDF A/1a	OpenDocument, OpenOffice 1.0, Rich Text Format 1.x, Office Open XML	MS-Word, Text602, 602 PC Suite, Amipro, WordPerfect
tabulky	Delimited text (CSV)	PDF OpenDocument, Office Open XML	MS-Excel, Calc602, Lotus
prezentace		PDF, OpenDocument, Office Open XML	MS-PowerPoint
rastrová grafika	TIFF, PNG	BMP, JPEG, JPEG2000, TIFF (komprimovaný LZW, JPEG), GIF	TIFF (jiná komprese), PCX, interní formáty grafických aplikací
vektorová grafika	SVG 1.1 (bez Javy)	Computer Graphic Metafile	interní formáty grafických aplikací
zvukové dokumenty	WAV, AIFF, Broadcast Wave	MP3, MP2, OGG Vorbis	Windows Media Audio, RealNetworks
video dokumenty	MPEG-1, MPEG-2, QuickTime, AVI (nekomprimované)	OGG Theora, MPEG-4	AVI, QuickTime (komprimované), Windows Media Video, RealNetworks

Tab. č. 1

Preferované formáty jsou vhodné pro dlouhodobé uchovávání a budou přebírány do archivu bez migrace. Akceptované formáty jsou pro dlouhodobé uchovávání méně vhodné, ale není nutno je v okamžiku předání migrovat do formátů preferovaných. Vyžadují však zvýšenou pozornost. Formáty ze skupiny neakceptovaných musí být nejpozději při předání do archivu migrovány nejlépe na preferovaný nebo alespoň na akceptovaný formát. Nejvhodnějším okamžikem pro převod dokumentu do preferovaného formátu se jeví okamžik jeho vyřízení (uzavření) u původce.

Stanovení preferovaných formátů však neznamená, že dokumenty musí vznikat pouze v nich. Dokument může být vytvářen v libovolném formátu, ale ve výše uvedeném okamžiku životního cyklu by měl být převeden do preferovaného formátu. Například u rastrové grafiky mohou použít formát GIMP či Photoshopu, který umožní plně využít vlastnosti editoru (např. vrstvy), ale poté je vhodný převod do PNG či TIFF. Podobně lze úřední dopis ukládat po dobu tvorby ve formátu MS-Office, a teprve v okamžiku uzavření (spisu apod.) ho převést do PDF/A.

V současné době jsou vládním usnesením č. 1338 z 3. listopadu 2008 stanoveny výstupní datové formáty statických dokumentů v digitální podobě ze systémů spisové služby vykonávaných elektronickou formou za použití výpočetní techniky a datové formát statických dokumentů v digitální podobě připravovaných pro předání do Národního digitálního archivu. Jedná se o formát PDF/A-1a (ISO 19005-1 – Portable Document Format – Electronic document file format for long-term preservation) pro statické textové, obrazové a kombinované dokumenty v digitální podobě a o formáty PNG (ISO/IEC 15948:2004 - Portable Network Graphics) a TIFF (Tagged Image File Format - revize 6 - nekomprimovaný) pro statické obrazové dokumenty v digitální podobě.

Autenticita

Autenticita je často skloňovaným problémem digitální archivace. Je nutno přiznat, že jde o dosud nevyřešený problém vyžadující širší veřejnou diskuzi a v současné době ho nelze jednoznačně rozhodnout. Projekt Národního digitálního archivu předpokládá, že po určitou dobu budou do archivu přicházet v nutných případech souběžně klasické i digitální archiválie a autenticitu elektronických dokumentů bude možno ověřit shodou s klasickými dokumenty.

V rámci technologického projektu je autenticita chápána jako prokázání, že dokument je tím, čím má být, že nebyl v průběhu uložení v digitálním archivu změněn a že procesem migrace nedošlo ke ztrátě žádné podstatné informace.

Zvoleným principem zachování autenticity v Národním digitálním archivu je fyzické a procesní zajištění uložení dokumentů oproti změně a transparentní, dokumentovaný způsob migrace. To vyžaduje certifikaci digitálního archivu jako důvěryhodného úložiště. Projekt předpokládá certifikaci archivu třetí stranou, že splňuje požadavky takového úložiště, a to jak po stránce fyzické bezpečnosti, tak po stránce procesní, včetně zajištění dostatečných finančních prostředků pro jeho dlouhodobou činnost. Certifikace proběhne podle aktuálních standardů či předpisů. Jednou z navržených možností je posouzení archivu Národním bezpečnostním úřadem.

Archiv bude považovat dokumenty předané původcem za autentické a zachová informace, jakými prostředky a s jakým výsledkem byla certifikace ověřena. Například u dokumentu podepsaném zaručeným elektronickým podpisem bude zachována informace, kdo dokument podepsal, že byl podpis ověřen a s jakým výsledkem. Dokument pak bude uložen v archivu spolu s těmito informacemi, ale v podstatě nepodepsaný. Autenticita dokumentu po uložení v archivu bude zajištěna vnitřními prostředky archivu.

Platný zaručený elektronický podpis (či elektronická značka) autora dokumentu (i s časovým razítkem odeslání) má svůj význam v prvotních fázích životního cyklu digitálního dokumentu, kdy po určité době potvrzuje autenticitu digitálního dokumentu. Jeho použití se jeví žádoucí maximálně do chvíle přijetí dokumentu do digitálního archivu. Je na původci, aby ověřil primární autentizační údaje o dokumentu a výsledek ověření uvedl do popisných metadat. Součástí metadat informačního balíčku je sekce pro vložení údajů o ověření elektronického podpisu. Projekt počítá s použitím elektronického podpisu na úrovni NDA pouze pro kontrolu integrity dat při přenosu od původce ke zpracování.

Metadata

Při dlouhodobém uchovávání digitálních dokumentů nestačí jen zajistit, aby byl dokument zachován a čitelný. K dokumentu je potřeba rovněž uchovat další informace, tzv. metadata. Metadata nám například pomohou zařadit dokument do kontextu doby jeho vzniku, poskytnou údaje o manipulaci s dokumentem, informují o formátu, v němž je dokument uložen apod.

V souladu s OAIS budou všechna metadata uložena spolu s dokumentem v rámci AIP (Archival Information Package).

Množinu metadat je možno rozdělit na tři hlavní skupiny: metadata popisná, konzervační (uchovávací) a strukturální. Při stanovení prvků metadat se vychází z mezinárodních norem a doporučení, které jsou v případě potřeby doplněny metadaty specifickými pro české prostředí. To se týká hlavně popisných metadat.

Popisná metadata (například název, popis, autor, původce, typ, kategorie, a další) slouží převážně pro vyjádření obsahu digitálních dokumentů a jsou využívána pro vyhledávání nebo zařazení objektu a k zjištění základních údajů o něm. Ačkoli mohou vznikat po celou dobu života dokumentu, vznikají hlavně ve dvou etapách životního cyklu: ve fázi příjmu a zpracování u původce a při archivním zpracování.

Fáze příjmu a zpracování u původce je pro vznik popisných metadat klíčová. Popisná metadata, která nebudou zachycena v této fázi životního cyklu, bude později velmi pracné, ne-li nemožné, doplnit. Je však třeba uvést, že záznam popisných metadat neznamená nárůst zátěže pracovníků, kteří s dokumenty pracují. Všechna metadata, která je potřeba zaznamenat, se již dnes k dokumentu běžně evidují (spisový znak, pořadové číslo, rok záznamu, odesílatel, příjemce apod.), případně je generuje systém správy dokumentů (ERMS) automaticky (systémový identifikátor, datum přijetí atd.). Z řečeného vyplývá, že systémy spisové služby a jim podobné informační systémy mají při vkládání popisných metadat nezastupitelnou roli.

Druhým okamžikem, ve kterém dojde k rozsáhlejšímu zadávání popisných metadat, je fáze archivního zpracování. Tato fáze není v současnosti příliš rozpracována, což je ale logické. Nejprve je potřeba dokumenty uchovat a teprve následně mohou být zpracovávány. Již dnes je však možno říci, že nutnost archivního zpracování je nepřímou úměrnou kvalitě popisu dokumentu u původce. Čím kvalitnější bude popis, o to menší bude potřeba archivního zpracování.

Popisná metadata vycházejí z doporučení Moreq2 vypracovaného organizací DLM fórum. Uvedené doporučení je předlohou připravovaného Národního standardu pro elektronické systémy spisové služby, který je adaptací Moreq2 na české prostředí a jehož součástí budou i definice metadatových prvků. Na úrovni archivního zpracování budou využívány standardy ISAD(G), ISAAR(CPF), EAD, EAC a množina popisných metadat bude rozšířena o prvky, které z těchto standardů vyplývají.

Konzervační metadata slouží pro podporu uchovávání a archivačních aktivit. Obsahují údaje o formátu souboru, ve kterém je dokument uložen, technické údaje o uložených digitálních objektech, informace o činnostech či změnách provedených s digitálním obsahem, o migracích dokumentu apod. Tato metadata budou rovněž vznikat po celou dobu životního cyklu dokumentu, převážně však v digitálním archivu. Je to logické, neboť digitální dokumenty trvalé povahy (digitální archiválie) budou větší část svého životního cyklu uloženy v archivu.

Uchování informace o formátu, ve kterém je dokument uložen, je zásadní pro případné znázornění dokumentu a pro volbu správného okamžiku a vhodného způsobu migrace. Informace o změnách dokumentu (vytvoření nového znázornění, migrace) a o manipulaci s ním nám poskytnou podporu při zajištění autenticity. Na základě důvěryhodnosti archivu a jeho schopnosti doložit veškeré operace s dokumentem pak bude možno usuzovat i na věrohodnost dokumentu.

Podobně jako u popisných metadat bude základ konzervačních metadat vycházet z Národního standardu pro elektronické systémy spisové služby. Tato základní množina pak bude rozšířena o metadata standardu PREMIS. K udržování informací o formátu souborů pak bude použit online registr formátů PRONOM.

Strukturální metadata slouží pro sdružení všech částí informačního balíčku do jednoho logického celku. Ukazují, jak spolu jednotlivé součásti dokumentu souvisí. Jak již bylo uvedeno, bude dokument v archivu uložen spolu se souvisejícími metadaty v rámci tzv. archivního informačního balíčku (AIP). Nelze však říci, že jeden AIP bude tvořen právě jedním dokumentem. AIP může být tvořen spisem a v takovém případě může obsahovat i několik dokumentů. Dokument sám pak nemusí být uložen v jednom souboru, ale může sestávat z několika souborů (komponent). Podobná situace je u metadat, která se mohou vztahovat k celému archivnímu balíčku, případně bude každý objekt (spis, dokument, komponenta) obsahovat svá vlastní metadata. Závislosti a vztahy jednotlivých částí popisují právě strukturální metadata.

Struktura informačního balíčku je definována normou OAIS. Pro jeho zápis bude využit standard METS a strukturální metadata budou vycházet právě z tohoto standardu.

Legislativa

Dosavadní legislativa péči o digitální dokumenty neřešila. V současnosti je připravována novela zákona č. 499/2004 Sb., o archivnictví a spisové službě, včetně prováděcích předpisů. V novele je mimo jiné navrženo, aby dokumenty v digitální podobě určené ke skartačnímu řízení musely být opatřeny metadaty podle národního standardu pro spisovou službu vykonávanou v elektronické podobě a dále v §45 odst. 3 se navrhuje, že Národní archiv na úseku péče o digitální dokumenty:

a) ukládá archiválie v digitální podobě vybrané k trvalému uložení veřejnými archivy s výjimkou archiválií vybraných bezpečnostními archivy zpravodajských služeb České republiky, udržuje jejich obsah a zajišťuje jejich trvalou čitelnost,

b) zpracovává a zpřístupňuje archiválie v digitální podobě převzaté od původců uvedených v odstavci 1 písm. a) a c) a archiválie v digitální podobě získané podle odstavce 1 písm. d),

c) plní metodickou a poradenskou funkci v oblasti předarchivní péče o dokumenty v digitální podobě, digitalizace dokumentů a archiválií.

V prováděcím předpise k tomuto zákonu pak budou určeny formáty těchto typů dokumentů předávaných do Národního digitálního archivu: PDF/A pro statické textové dokumenty, PNG a TIFF pro statické obrazové dokumenty, WAV pro zvukové dokumenty a MPEG-2 pro audiovizuální dokumenty.

Závěr: Aktuální stav a plánovaný postup

Jak již bylo zmíněno, bude projekt Národního digitálního archivu na základě vládního usnesení spolufinancován ze strukturálních fondů EU. Aktuálně je připravena velká část podkladů pro projektovou žádost. Po vyhlášení výzvy pro osu 1 Integrovaného operačního programu bude podána projektová žádost a po jejím případném schválení bude projekt zahájen.

Oproti technologickému projektu není budování digitálního archivu pojímáno jako jedna zakázka, ale je rozděleno na stavební a technologickou část. Důvodem tohoto rozdělení je skutečnost, že jde o velmi odlišné problematiky. Stavební část zahrnuje vypracování projektové dokumentace a vlastní stavbu objektu pro hlavní pracoviště a rekonstrukci stávajícího objektu pro záložní pracoviště. Součástí stavební části je i tzv. technologické vybavení budov (EPS, EZS, klimatizace apod.). Technologická část pak zahrnuje dodávku hardware, software a vytvoření informačního systému archivu.

V případě, že se naplní předpoklady, měla by být v červnu 2010 hotova rekonstrukce záložního pracoviště a v únoru 2011 zkolaudován objekt hlavního pracoviště. V dubnu 2011 by měla být instalována technologická část tak, aby mohl v létě 2011 proběhnout zkušební provoz, na který naváže audit a certifikace archivu ukončená v listopadu 2011.

1. ledna 2012 by byl zahájen rutinní provoz Národního digitálního archivu.