2.3 Technologické podmínky pro vybudování důvěryhodného archivu

2.3.1 Obecné podmínky

Při posuzování vlivu vlastností informačního systému na důvěryhodnost digitálního archivu jako celku lze vycházet ze základního rozdělení vlastností na:

- Funkční vlastnosti
  Jedná se o ty vlastnosti, které plní funkce z hlediska procesů archivu. Funkční požadavky vychází z politiky a postupů archivu, které zase vychází z životního cyklu digitálního dokumentu tak, jak je popsán v kapitole □. Z hlediska funkčních vlastností je především důležité, aby tyto byly ve shodě s procesy archivu a aby poskytovaly podporu procesům archivu. Přestože tento požadavek se jeví jako samozřejmý, v praxi není často dodržen a řada procesů je naopak přizpůsobena vlastnostem dostupných systémů, a tak řada kroků v rámci procesu je pak buď nevyhovující nebo nadbytečná. Takový systém tedy nepřímo přispívá k degradaci procesů v organizaci archivu, což má samozřejmě i negativní důsledky na důvěryhodnost. Při výběru systémů archivu je tedy nutné věnovat pozornost detailní specifikaci funkčních požadavků a vyhodnocení shody systému s nimi. Ne všechny požadavky mohou být známé nebo dobře definované na počátku, proto je nutné věnovat zvýšenou pozornost otevřené a flexibilní architektuře systému, která umožní rozšiřování funkčnosti podle potřeb – detailnější rozbor takové architektury je uveden v kapitole 2.3.6.
- Obecné (nefunkční) vlastnosti
  Jsou obecné vlastnosti informačních systémů, společné většině implementací, které celkově ovlivňují chování systému. Tato kapitola se dále bude zabývat právě vlivem vybraných obecných nefunkčních vlastností na digitální archiv.

Mezi hlavní nefunkční vlastnosti, důležité pro posuzování systému digitálního archivu, patří (seřazeny podle důležitosti):

- Vysoká spolehlivost a dostupnost – jsou klíčovými vlastnostmi archivního systému, kdy systém musí zajistit prakticky neomezený přístup k uloženým dokumentů po požadovanou dobu, která je řádově v desítkách let, případně i neomezená. Redundance a zdvojování systémů je typickým prvkem, který je používán pro zajištění těchto vlastností.
- Bezpečnost – je další klíčovou vlastností, kdy systém musí zajistit především řízení přístupu jak k uloženým dokumentům, tak k vlastním funkcím systému. Velmi důležitou součástí bezpečnosti je i zajištění integrity dat uložených v systému archivu, což se týká jak samotných elektronických dokumentů, tak metadat a indexových databází. Pro určité typy dat může být vyžadováno i zajištění jejich důvěrnosti.
- Škálovatelnost – tedy schopnost systému růst a přizpůsobovat se rostoucímu objemu uložených dat, počtu přijatých dokumentů pro archivaci a počtu požadavků na vyhledání uložených dokumentů a přitom zajistit stabilní výkonnost systému. Je zřejmé, že škálovatelnost patří k základním vlastnostem digitálního archivního systému, kde nárůst objemu zpracovávaných dat má často exponenciální tendenci.
- Účtovatelnost/Auditovalenost – znamená schopnost systému zaznamenávat a uchovávat události a akce v systému. Podrobné záznamy o historii dokumentů

jsou velmi důležitým faktorem pro podporu jejich důvěryhodnosti, většina norem a doporučení z této oblasti (viz následující odstavec) specifikuje, které auditní záznamy je potřebné zaznamenat.

- Snadná spravovatelnost – komplexní systém jako digitální archiv vyžaduje řadu úkolů souvisejících se správou systému samotného. Jejich rychlé a efektivní vykonání podmiňuje funkce systému samotného.

Výše jmenované obecné vlastnosti systému tvoří pouze rámcové východisko pro posuzování podmínek důvěryhodnosti v systému digitálního archivu, podrobnější specifikace jak vlastností informačního systému, tak navazujících procesů je specifikována v řadě mezinárodních standardů a doporučení, které lze využít pro detailnější specifikace systémů digitálního archivu. Relevantní dokumenty z této oblasti lze rozdělit do tří oblastí:

- Normy a doporučení týkající se dlouhodobé archivace nebo uchování elektronických dokumentů (sem patří například ISO/TR 15801[38], ISO/TR 18492[39] a RLG doporučení pro důvěryhodná digitální úložiště[40])
- Normy a doporučení týkající se systémů správy záznamů (Record Management), kdy archivní systémy mají v řadě aspektů blízko k těmto systémům (jedná se např. o Evropské doporučení Moreq[41] nebo ISO 18459[42])
- Normy a doporučení týkající se bezpečnosti IS (jako je např. ISO/IEC 17799[43])

Kromě obecných podmínek popsaných výše, které jsou známy i z řady jiných systémů, musí systém archivu používat řadu dalších specifických technologií a postupů, které jsou úzce zaměřeny na dlouhodobé uchování elektronických dokumentů. Tyto specifické technologické vlastnosti jsou popsány v dalších kapitolách.

2.3.2 Uchování nezávislé využitelnosti formátu elektronického dokumentu

Při výběru vhodných formátů elektronických dokumentů je třeba vzít do úvahy dvě skupiny faktorů:

- koncepční faktory, které ovlivňují trvalost jakéhokoliv digitálního formátu
- faktory vztahující se ke kvalitě nebo speciální funkcionalitě, která může být požadována pro určité kategorie obsahu

V současné době existují stovky různých formátů z nichž řada má verze a podverze. Obecně užívané názvy formátů nenabízí dostatečné rozlišení pro archivační účely. Názvy formátů, včetně koncovek souborů jako jpg, pdf, mov, jsou příliš generické a

[38] ISO/TR 15801:2004 Electronic imaging -- Information stored electronically -- Recommendations for trustworthiness and reliability

[39] ISO/TR 18492:2005 Long-term preservation of electronic document-based information

[40] Research Libraries Group. (2002). Trusted digital repositories: Attributes and responsibilities. An RLG-OCLC Report. Available at: <http://www.rlg.org/longterm/repositories.pdf>

[41] MoReq SPECIFICATION - MODEL REQUIREMENTS FOR THE MANAGEMENT OF ELECTRONIC RECORDS, IDA Programme, CECA-CEE-CEEA, Bruxelles- Luxembourg, 2001

[42] ISO 15489-1:2001Information and documentation -- Records management -- Part 1: General
ISO/TR 15489-2:2001Information and documentation -- Records management -- Part 2: Guidelines

[43] ISO/IEC 17799, Information technology-Security techniques-Code of Practice

neumožňují rozlišit různé typy a verze. Jako poměrně jednoduchý příklad může posloužit formát TIFF. TIFF může obsahovat bitmapy v různých zakódováních - nekomprimovaných, komprimovaných s použitím bezztrátového LZW algoritmu nebo pro bitonální obrazy komprimované s použitím ITU G4 komprese. Případná budoucí transformace nebo migrace se bude odvíjet od konkrétně použitého zakódování - pro G4 TIFF bude pravděpodobně vhodné vybrat jiný cílový formát než pro nekomprimovaný 24-bitový TIFF. TIFF má dále různé verse odrážející požadavky na kvalitu a funkcionalitu - TIFF/EP (ISO standard pro digitální fotografii) a TIFF/IT (ISO standard pro předtiskové předlohy). Jako další komplexnější příklad mohou posloužit soubory s koncovkou .pdf. PDF formát obecně vychází z Adobe Portable Document Format. Adobe tento formát postupně zdokonalovalo a přidávalo do něj další vlastnosti, což se odráží v různých PDF verzích. PDF lze nyní použít jako poměrně jednoduchý formát pro stránkovaný text, ale též jako obálku pro mnoho různých obrazových formátů (TIFF, JPEG, JPEG2000) nebo „kontejner“ pro komplexní dokumenty a interaktivní multimediální soubory. Kromě toho původní PDF vedlo k vytvoření dalších standardů - PDF/X (ISO standard 15930 pro předtiskové sestavy) a PDF/A (ISO standard 19005 pro dlouhodobou archivaci).

Tato skutečnost je obecně archivační komunitou reflektována a vede k vytváření registrů digitálních formátů, v nichž je možné nalézt detailní popisy formátů a jejich verzí. Jako dobré referenční příklady jmenujme Global Digital Format Registry (GDPR) (projekt Harvardské University) - http://hul.harvard.edu/gdfr/, nebo PRONOM (online registr formátů a software produktů, tyto formáty podporujících spravovaný britským Národním archivem) - http://www.nationalarchives.gov.uk/pronom/.

Správce digitálního archivu bude muset mít dobré znalosti těchto formátů jejich variant a versí. TIFF

2.3.2.1 Faktory ovlivňující spravovatelnost (sustainability) formátů

Literatura identifikuje sedm faktorů, které mají vliv na dlouhodobé uchování dokumentů a náklady s tímto spojené. Tyto faktory se ukáží jako významný též v budoucnu při migraci na nové formáty, emulaci dnešního software na budoucích počítačích anebo hybridních strategiích.

Otevřenost formátu (disclosure)

Uchování obsahu v daném formátu není možné bez detailní znalosti toho, jak jsou ve formátu ukládány informace na úrovni jednotlivých bitů a bytů. Tato znalost je velmi důležitá pro možnost ověření integrity formátu ukládaného dokumentu. Pro neproprietární, otevřené formáty jsou většinou jak podrobná dokumentace formátu, tak validační nástroje lépe dostupné než pro formáty proprietární.

Příklady:

- TIFF, dobře popsaný, mnoho nástrojů od třetích stran;
- MrSID, proprietární GIS komprimovaný formát (LizardTech), pouze částečně dokumentovaný;
- JPEG2000 Part 1, otevřený standard, zcela zdokumentovaný.

Rozšířenost (Adoption)

Odkazuje na stupeň používání formátu zainteresovanými uživateli IT a tvůrci obsahu. Značná rozšířenost formátu snižuje pravděpodobnost jeho rychlého zastarání. Pro rozšířený formát je větší pravděpodobnost vzniku migračních a emulačních nástrojů v rámci IT průmyslu, bez nutnosti, aby se archivní instituce na vzniku těchto nástrojů přímo podílela. Dobrými kritérii rozšířenosti formátu je standardní dodávka nástrojů podporujících daný formát přímo s PC, nativní podpora formátu přímo ve Webových prohlížečích a/nebo existence řady konkurenčních produktů pro tvorbu obsahu, manipulaci s obsahem anebo získávání obsahu z daného formátu. Podpora formátu jinými významnými archivními institucemi má též svoji váhu.

Transparentnost

Faktor popisující přístupnost obsahu uloženého v daném formátu pro přímou analýzu s použitím základních nástrojů, včetně čitelnosti lidským okem v jednoduchém textovém editoru. Digitální formáty, v nichž je relevantní informace uložena přímo a jednoduše, lze snadněji migrovat, psát pro ně odpovídající prohlížeče a v budoucnu jsou též vhodnější pro „digitální archeologii“.
Čitelnost textových dokumentů je příznivě ovlivněna použitím standardního znakového kódování (např. UNICODE s použitím UTF-8 kódování) a ukládáním v posloupnosti odpovídající přirozenému čtení. Tato poznámka platí i pro metadata vkládaná do textových i netextových souborů. Pro ukládání počítačových programů je zdrojový kód podstatně transparentnější než kompilovaný. Co se týká netextových informací základní reprezentace jsou transparentnější než reprezentace optimalizované pro úspornější ukládání, rychlejší zpracování, atd. Jako příklad základní reprezentace uveďme uložení rastrové grafiky jako nekomprimovanou bitovou mapu.
Šifrování je samozřejmě v přímém rozporu s transparentnostní, komprimace též transparentnost potlačuje. Na druhou stranu, z řady praktických důvodů, řada audiálních, obrazových a video digitálních děl nebude nikdy ukládána v nekomprimovaném formátu a to ani v okamžiku vzniku. Digitální archivy samozřejmě budou chtít přijímat tato díla; jedním z kriterií vhodnosti formátu je pak použití otevřeného, zdokumentovaného a široce rozšířeného kompresního algoritmu.
Příklady:

- TIFF, nekomprimovaný, přímočaré kódování, v případě ztráty specifice si lze představit úspěšné použití reverzního inženýrství pro přečtení;
- JPEG2000, part 1, komplexní komprimované kódování, další faktory, jako např. velká rozšířenost snižují možnost, že kompresní algoritmus upadne v celkové zapomenutí.

Sebedokumentace (Self-documentation)

Proces dlouhodobé archivace digitálních objektů se výrazně ulehčuje pokud součástí digitálního objektu jsou základní deskriptivní metadata (podobně jako titulní strana knihy), data technické a administrativní povahy zachycující vznik dokumentu a předchozí stavy jeho životního cyklu. Všechny tyto informace umožňují a ulehčují správné zobrazení a pochopení objektu a z hlediska dlouhodobého uchování dokumentu je mnohem lepší, pokud jsou tyto informace uloženy přímo v daném objektu a ne někde odděleně.

V souvislosti s vytvářením a sdílením digitálního obsahu potřeba formátů s dobrými možnostmi pro ukládání metadat rostla. Tato potřeba ovlivnila definice nových verzí stávajících formátů a vznik formátů nových. Jako ilustrativní příklad uveďme původní JPEG standard, který obsahuje podporu pro velmi skromná metadata, později rozšířený o EXIF JPEG, který kombinuje původní JPEG kompresi s bohatšími metadatami. Nakonec byl vytvořen nový JPEG2000 standard. Part 2 JPEG2000 standardu podporuje DIG35 metadatová schémata a umožňuje vložení libovolných metadat do tzv. metadatových boxů.

Bohatá a dobře definovaná metadata umožňují digitálnímu archivu tato data v průběhu procesu ukládání extrahovat a uložit odděleně do metadatových úložišť anebo archivních katalogů a tak zefektivnit proces zpětného vyhledávání dokumentů.

Referenční model OASIS zdůrazňuje potřebu metadat, která poskytnou o uloženém digitálním objektu následující informace:

- Reprezentace (umožní zobrazení dat a prohlížení jako smysluplné informace);
- Reference (identifikace a popis obsahu);
- Kontext (např. důvod pro vytvoření dokumentu);
- Stabilita (informace umožňující ověření integrity obsahu);
- Původ (data zachycující vytvoření a vlastnictví dokumentu).

Vnější závislosti

Faktor popisující stupeň závislosti daného formátu na specifickém hardware, operačním systému, prohlížecím software, včetně prognózy možností zvládnutí těchto závislostí v budoucnu. Jako ilustraci uveďme příklady interaktivního obsahu, který je navržen tak, aby pracoval se speciálním hardware – joystick. Dalším příkladem jsou vědecká data sbíraná za pomocí datových senzorů, která mohou potřebovat speciální software pro analýzu a vizualizaci.

Vliv patentů

Možnosti a náklady archivních organizací uchovávat obsah v určitém formátu jsou zásadně ovlivněny patenty s formátem souvisejícími. Existence patentů zpomaluje vývoj open-source software pracujících s daným formátem a ovlivňuje cenu komerčního software. Je dobré si uvědomit, že problém není v existenci patentu, ale v podmínkách, které mohou držitelé patentu uplatnit. Pokud licenční ujednání zahrnuje poplatky odvozené od užívanosti, náklady spojené s dlouhodobým uchováním dokumentu v tomto formítu mohou být vysoké a nepředvídatelné.
Existence patentů a související licenční politika mohou výrazně ovlivnit rozšířenost formátu, jak ukazuje příklad s ISO formátem MPEG. MPEG-1 nevyžaduje žádné licencování. MPEG-2 licence nutí výrobce zařízení k licencování technologie a tak každé zařízení, které umožňuje tvorbu MPEG-2 formátu, je zatížené poplatkem. MPEG-4 jde ještě dál a uplatňuje „pay-per-view“ licenční politiku. To vedlo k tomu, že MPEG-4 se prakticky mezi výrobci neujal a vznikly jiné podobné, ale nestandardizované formáty.

Ochranné mechanismy snižující použitelnost digitálního objektu

K efektivní správě digitálního obsahu a k poskytování přiměřených služeb budoucím uživatelům musí mít správce archivu možnost replikovat uložený obsah na nová média, migrovat anebo jinak normalizovat tak, aby obsah byl odpovídajícím způsobem využitelný v budoucnu při použití nových technologií. Dlouhodobé uchování obsahu je obtížné anebo může být přímo znemožněné při aplikaci některých ochranných mechanismů.
Příkladem takových mechanismů jsou formáty svázané s určitými fyzickými nosiči nebo zařízeními. Některé formáty mají zabudované mechanismy (časová omezenost, vazba na specifické hardware, aktivní síťové připojení) omezující použitelnost kvůli ochranně IP.

2.3.2.2 Faktory kvality a funkcionality

Kvalita a funkcionalita jsou faktory svázané se schopností formátu reprezentovat podstatné vlastnosti daného obsahu tak, jak je požadováno nebo očekáváno současnými nebo budoucími uživateli. Tyto faktory se mění podle žánru a v daném žánru též v závislosti na formě. Podstatné charakteristiky jsou jiné pro zvuk a obraz a ne každý obrazový formát je vhodný pro fotografii.
Vzhledem ke značné rozmanitosti je dobré se pro začátek soustředit na čtyři typické obsahové kategorie: statický obraz, video, text a zvuk. Dalším krokem je pak ukládání exotičtějšího obsahu jako např. WEBových stránek nebo databází. Jak ukazují současné pokusy a výzkumy databáze bude nutné kategorizovat podle obsahu – geospaciální data, sociální průzkumy, apod.
Literatura[44] navrhuje pracovat s konceptem normálního prohlížení (normal rendering) jako se základem pro prezentaci obsahu uživateli (statický obraz může být zvětšován, zvuk lze přehrát, zastavit).
Následující seznam rekapituluje kvalitativní a funkční faktory ve vazbě na formáty pro statické obrazy:

- Normální prohlížení zahrnuje prohlížení na obrazovce a možnost zvětšení. Je podporován standardní tisk;
- Rozlišitelnost – stupeň v jakém lze reprezentovat obsah s velkým rozlišením v daném formátu. U bitových obrázků je kvalita korelována s počtem pixelů na jednotku plochy a s bitovou hloubkou;
- Barevná věrnost – míra možnosti řídit barevný gamut reprezentovaný v daném obraze;
- Podpora pro grafické efekty a typografii – např. podpora vrstev.

[44] Caroline Arms, Carl Fleischhauer - Digital Formats: Factors for Sustainability, Functionality, and Quality, IS&T Archiving 2005 Conference, Washington, D.C.

2.3.2.3 Výběr formátů v praxi - porovnávání faktorů

V praxi znamená výběr vhodných formátů pro dlouhodobé uchovávání najít rovnováhu mezi výše uvedenými faktory, tj. ve stručnosti mezi spravovatelností, kvalitou a funkcionalitou. Jeden ze způsobů jak tuto komplexní otázku řešit je vytvoření a použití vyhodnocovací tabulky. Níže uvádíme ilustrativní případ použití pro ohodnocení některých formátů pro bitmapové obrázky (postup vyhodnocení je pouze naznačen, konkrétní hodnocení samozřejmě závisí od přesných požadavků archivu).

	TIFF	EXIFF-TIFF	JPEG	JP2000	MrSID
Otevřenost	+	+	+	+	.
Rozšířenost	+	+	+	.	.
Transparentnost					
Sebedokumentace					
Vnější závislosti					
Vliv patentů					
Ochranné mechanismy					
Rozlišení					
Barevná věrnost					

2.3.3 Metadata

Metadata („data o datech“) jsou dodatečné informace o elektronických dokumentech, získaná buď od původce, z vlastností a obsahu dokumentu nebo dodaná archivem. Každý dokument má svoje metadata, která jsou s ním pevně svázána.
Metadata jsou jak technicky, tak procesně jedním z nejdůležitějších aspektů archivace, jejich význam spočívá především v:

- vyhledávání elektronických dokumentů v archivu
- reprezentaci a zobrazení elektronických dokumentů
- prokazování autentičnosti a důvěryhodnosti elektronických dokumentů
- správě elektronických dokumentů.

Problémem metadat pro archivaci se zabývá řada národních a mezinárodních projektů a standardů. Na prvním místě definují standardy sémantiku metadat pro popis elektronických dokumentů a speciálně pro jejich archivaci – tedy jednotlivé informační prvky, jejich význam a obsah. Mezi hlavní standardy v této oblasti, které slouží často jako východisko pro ostatní standardy a pro konkrétní implementace, patří především:

- Dublin Core[45] - tento standard popisuje základní sadu metadat (18 základních prvků), popisující libovolný informační zdroj referovaný elektronicky. Vzhledem ke svému obecnému zaměření, kompaktní sadě prvků a široké podpoře je Dublin Core také často používán jako základ pro metadata

[45] Dublin Core Metadata Initiative, http://dublincore.org/, též ISO 15836

v archivech, kde jsou jeho základní prvky rozšířeny o další prvky, specifické pro dlouhodobé uchování elektronických dokumentů.

- OAIS[46] –popisuje obecný rámec pro archivační systém, kde definuje také informační model pro metadata příslušná k archivovanému elektronickému dokumentu.
- PREMIS[47] - je doporučení, vypracované mezinárodními organizacemi zabývajícími se výzkumem v oblasti knihovnictví a archivnictví, které popisuje podrobný návrh metadat (slovník prvků) pro elektronickou archivaci. Navržená struktura prvků vychází z OAIS modelu (ale také ze zkušeností řady konkrétních projektů) a rozvádí jej do konkrétních detailů.

Na výše uvedené pak navazuje řada národních a mezinárodních standardů a doporučení, které je rozvádějí pro konkrétní podmínky. Také řada norem a doporučení vznikala současně nezávisle jedna od druhé, takže existuje poměrně rozsáhlá dokumentace v této oblasti, kdy může být často obtížné zhodnotit přesný význam jednotlivých dokumentů. Standardizace v této oblasti není tedy zdaleka ukončena a stále se vyvíjí.

Metadata pro elektronické dokumenty určené k uschování by měla obsahovat především:

Jednoznačný, dlouhodobý identifikátor archivovaného dokumentu (resp. archivního balíku)

Jednoznačná identifikace slouží pro referenci a vyhledání balíku a v něm obsaženém dokumentu. Identifikátor musí být jednoznačný po celou dobu uložení balíku pro všechny možné balíky v rámci archivu, případně řady archivů (proto se také volí většinou globálně jednoznačný identifikátor, v řadě aktuálních doporučení se vychází např. z URI[48]).

Zajímavým aspektem je také zabezpečení jednoznačného identifikátoru, kdy je vhodné, aby tento nemohl být snadno zaměněn identifikátorem na jiný objekt. Proto některá schémata jako součást identifikátoru používají otisk(haš) dokumentu, na který odkazují (z hlediska dlouhodobého uložení je zde ale nutno také zohlednit životnost hašovacího algoritmu).

Popis obsahu elektronického dokumentu

Typicky zahrnuje několik prvků metadat, které popisují obsah uloženého elektronického dokumentu, slouží především pro pozdější vyhledávání dokumentu a kategorizaci dokumentů uložených v archivu. Patří sem prvky jako stručný popis, obsah, titul, autor, kategorie, klíčová slova aj.

Popis formátu elektronického dokumentu (reprezentace dat)

[46] CCSDS 650.0-R-2. 2001, Reference Model for an Open Archival Information System (OAIS); též jako ISO 14721:2003

[47] Data Dictionary for Preservation Metadata, Final Report of the PREMIS Working Group, May 2005, Preservation Metadata: Implementation Strategies (PREMIS), A working group jointly sponsored by OCLC and RLG, http://www.oclc.org/research/projects/pmwg/default.htm

[48] BERNERS-LEE, T., FIELDING, R., AND MASINTER, L. 1998. Uniform Resource Identifiers (URI): Generic Syntax. IETF RFC 2396.

Pro pozdější využití elektronického dokumentu je klíčové detailně v metadatech popsat formát dokumentu a způsob jak má být reprezentován do podoby vhodné pro lidské vnímání. Popis by měl zahrnovat (jméno, verzi, aj.):

- Vlastní formát dat (např. PDF ver. 1.4)
- Aplikaci, která vytvořila dokument nebo je určená pro reprezentaci dat (např. Adobe Reader 7.0.8)
- Operační systém, pro který je aplikace určena (např. MS Windows XP ver. 2002 SP.2)
- Referenční platformu – popis počítače, na kterém daný operační systém a aplikace zobrazí data (např. PC s procesorem Pentium, 256MB paměti)

Reference na externí zdroje

Metadata mohou obsahovat reference (odkazy) na různé elektronické dokumenty a jiné relevantní zdroje informací. Reference mohou odkazovat jak na jiné archivní balíky v rámci jednoho archivu nebo i jiné relevantní externí informace (zde je nutné vzít v potaz dostupnost a „životnost“ odkazů tak, aby byly platné pokud možno po dobu uložení dokumentu).

Reference mohou odkazovat na např.:

- jiné verze dokumentu
- odkazy na osoby a organizace spojené s uloženým dokumentem
- detaily k formátu a jeho reprezentaci – detailní popis formátu a způsobu jeho reprezentace do podoby vhodné pro lidské vnímaní
- různé detaily týkající se podmínek využití dokumentu
- a cokoliv dalšího, co může být zajímavé pro komunitu uživatelů v souvislosti s daným dokumentem.

Vzhledem k rozsáhlému množství typů referencí, reference kromě vlastního odkazu musí obsahovat i přesný kvalifikátor typu a významu odkazu.

Kontext elektronického dokumentu

Uvádí informace o způsobu vytvoření dokumentu, kdo jej vytvořil a za jakým účelem.

Historie elektronického dokumentu

Zde jsou zaznamenané události, které jsou významné pro historii dokumentu z hlediska archivace, týkají se:

- převzetí
- uložení

- migrace medií, migrace formátu atd.

Pro jednotlivé události z historie je důležité zaznamenat jejich přesnou specifikaci, datum a čas a identifikaci subjektů, které se na dané události podíleli.

Ochrana elektronického dokumentu

Obsahuje detaily týkající se ochrany integrity a autentičnosti obsahu elektronického dokumentu i celého archivačního balíku (tedy případně i příslušných metadat). Přesné informace uložené v metadatech záleží především na způsobu zabezpečení dokumentu. Je důležité uvést zde všechny detaily, které umožní ověřit zabezpečení během celé doby uložení dokumentu (viz také detailní popis zabezpečení elektronických dokumentů v kapitole 2.3.5).

Řízení přístupu k elektronickému dokumentu

Metadata, která slouží pro řízení přístupu k uloženému dokumentu v archivu. Požadavky na kontrolu a omezení přístupu jsou uplatňovány systémem archivu jak směrem k personálu archivu, tak ke komunitě uživatelů a vyplývají buď z obecných společensko-právních norem platných pro dokument, politiky archivu nebo z požadavků původce dokumentu.

Autorská práva

Informace týkají se autorských práv k uloženému elektronickému dokumentu a případných omezení v přístupu k dokumentu vyplývající z autorských práv.

Dalším důležitým aspektem pro implementaci metadat je způsob jejich kódování, tj. jak jsou převedena do posloupnosti bytů, která může být uložena v archivu spolu s elektronickým dokumentem. Tento aspekt je neméně důležitý, neboť je nutné zajistit, že metadata budou přístupná, čitelná a srozumitelná po celou dobu uložení elektronického dokumentu.

Kódování metadat musí vycházet z pevných standardů, které musí být přesně popsány a jejich popis udržován v rámci archivu. Kódování by mělo být nezávislé na technologii a umožnit tak přechod archivu na nové technologie bez nutnosti změn v metadatech. V současné době se jeví jako vhodným způsobem kódování metadat s využitím standardů založených na XML[49]. Např. METS[50] je jedním z často uváděných standardů z této oblasti, tento poskytuje jakýsi kontejner, v rámci kterého mohou být uloženy metadata z různých dalších standardů (jako je Dublin Core nebo PREMIS uvedené výše), které definují detailně sémantiku jednotlivých prvků. Pro tyto standardy a jejich konkrétní implementace METS definuje takzvané profily. Kromě toho METS také definuje strukturální informace a umožní popsat dokument složený z různých částí a také definovat odkazy na fyzické uložení jednotlivých částí. METS definuje XML schéma, které obsahuje 6 částí, kde mohou být umístěna jednotlivá metadata:

[49] Extensible Markup Language (XML) 1.0 (Fourth Edition) W3C Recommendation 16 August 2006, Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, François Yergeau eds.

[50] Metadata Encoding and Trasmission Standard, Library of Congress, http://www.loc.gov/standards/mets/

- hlavičku <metsHdr>
- popisná metadata <dmdSec>
- administrativní metadata <amdSec>
- skupiny objektů <fileSec>
- strukturální mapa <structMap>
- chování podle obsahu <behaviourSec>

Dále je také důležité vlastní uložení metadat, kdy je nutné stanovit kde a jakým způsobem budou data uložena v archivu a jak bude zajištěno jejich uchování. Uložení metadat je úzce spjato s jejich kódováním, kdy určité kódování předurčuje vhodné uložení metadat. V současných implementacích jsou využívány především tyto způsoby uložení:

- Jako tabulky v relační databázi
  Výhodou je zde především snadné vyhledávání v metadatech a jejich jednoduchá aktualizace.
- Jako strukturovaná data (typicky XML) uložená v XML nebo objektových databázích
  Výhody jsou obdobné jako u relačních databází navíc tyto technologie umožňují lépe zachovat vysokou strukturovanost a rozsáhlou hierarchii metadat. Nicméně tyto technologie nejsou zase prověřené pro velké objemy dat a výkonnost a spolehlivost zde mohou být problematické.
- Jako XML soubory uložené společně s elektronickými dokumenty, kdy spolu tvoří jeden archivní balík
  Velkou výhodou u tohoto přístupu je těsné spojení metadat s uloženým objektem, takže je těžší je oddělit, což je pozitivní z hlediska důvěryhodnosti a autentičnosti. Také se na takto uložená metadata dají použít stejné strategie dlouhodobého uložení a lze tak lépe zajistit trvání metadat po dobu uložení dokumentu.
- Textový soubor privátního formátu (tzv. flat file) uložený s dokumentem
  Vzhledem k široké podpoře XML a jeho standardizaci je tento přístup implementován spíše u starších řešení nebo tam, kde je nutná interoperabilita s nějakými staršími systémy.

Ukazuje se, že v praxi je nejvhodnější kombinace několika způsobů uložení metadat – jednak spolu s elektronickým dokumentem jako XML (kde budou všechna metada spolu s dokumentem) a jednak v databázi (kde bude možné podle metadat vyhledávat, nemusí zde být tedy všechna, jen ta která jsou využita pro vyhledávání nebo která vyžadují častou aktualizaci).

Pokud jsou data uložena společně s dokumentem je nutné také použít vhodnou obálku, která je spojí dohromady do jednoho balíku, identifikuje jednotlivé části a umožní přístup k nim a jejich pozdější využití. Standard METS může plnit i tuto funkci, kdy XML vytváří tuto obálku.

Je zřejmé, že metadata, jejich kódování a také obálka, která je spojuje s vlastním elektronickým dokumentem, mohou být různé během rozdílných etap životního cyklu elektronického archivu. Standard OAIS proto také definuje tři základní typy metadat (respektive typy balíku metadat + elektronický dokument):

1) SIP (Submission Information Package) – Předávací balík
   Označuje balík metadat a dokumentů tak, jak jsou předávány původcem do archivu.
2) AIP (Archival Information Package) – Archivní balík
   Označuje balík metadat a dokumentů tak, jak je uložen v archivu po jejich převzetí, ověření a případném doplnění.
3) DIP (Dissemination Information Package) - Výstupní balík
   Označuje balík metadat a dokumentů tak, jak je předán uživateli, který požaduje určitý dokument z archivu, kdy metadata budou obsahovat pouze informace relevantní pro uživatele, taktéž formát/kódování, obálka aj. musí být podřízeny možnostem uživatele.

Archiv pro každý z těchto typů musí definovat pevná pravidla, jak pro obsah metadat, tak i kódování, uložení, obálky, přenosové protokoly atd., kde některé části mohou být společné (např. slovníky metadat, XML kódování aj.). Pro převody mezi jednotlivými typy musí existovat dobře definované procedury.

2.3.4 Média pro uložení elektronických dokumentů

Pro dlouhodobé uchování elektronických dokumentů je nutné zajistit jejich uložení na vhodných nosičích – mediích a náležitou péči o tyto média po celou dobu uchování elektronických dokumentů.

Z tohoto hlediska je nutné především zohlednit následující aspekty:

- Architektura systému úložiště dat
- Výběr vhodného typu medií
- Proces kontroly a údržby medií během uchování
- Zestárnutí technologie úložiště dat a přechod na nový systém

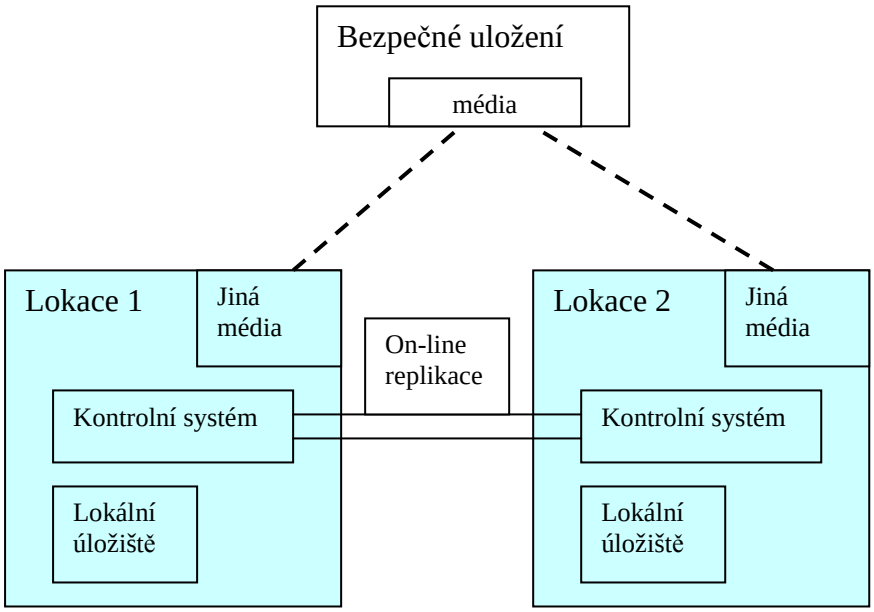
2.3.4.1 Architektura systému úložiště dat

Přestože vlastnosti nosičů jsou důležité pro uchování dat, důležitější je návrh vhodné architektury systému úložiště dat, který by měl vycházet z detailní analýzy požadavků na uchování dokumentů. Architektura také bude vodítkem pro výběr vhodného typu medií.

Jedním z hlavních principů, který utváří architekturu úložiště dat pro archiv, je princip redundance. Z dlouhodobé perspektivy nelze žádnou technologii považovat za bezchybnou a také je nutné počítat s událostmi typu přírodní katastrofy a podobně. Proto musí architektura umožnit nezávislé uložení dat na několika nezávislých technologiích a nezávislých lokacích zároveň.

Tyto redundance pak musí být propojeny do jednoho funkčního systému tak, aby při výpadku jedné části, mohly být služby poskytovány ostatními částmi systému (při zajištění řízeného a pokud možno plynulého přechodu).

Typickým příkladem takovéto architektury může být systém se dvěmi nezávislými středisky, která pracují v on-line módu s třetí off-line zálohou, která je držena na jiném bezpečném místě, viz následující obrázek:



Obdobná architektura[51] tvoří základ řady moderních digitálních archivů.

Současné výzkumy přichází také s dalšími možnostmi, které využívají sítě Internet. Redundance je zde o řád až dva vyšší než v předchozím řešení (tj. jsou uchovávány až stovky kopií jednoho elektronického dokumentu) s tím, že je pak požadována poměrně nízká spolehlivost (i případná důvěryhodnost) jednotlivých úložišť. Takový systém pak typicky zahrnuje řadu organizací, které se na jeho provozu podílí. Příkladem může být třeba systém LOCKSS[52].

Kromě redundance na úrovni replikace úložišť by měla existovat i redundance na úrovni práce s jednotlivými bloky dat až po úroveň jednotlivých bytů, především za využití kontrolních součtů (CRC) a binárních samoopravných kódů. Veškerá uložení a přesuny elektronických dokumentů v archivu by měli být zajištěny pomocí těchto mechanismů, které pak podpoří vlastnosti digitálních medií a přenosových kanálů.

Dalším důležitým kriteriem pro architekturu úložiště je jeho rozdílné využití v systému archivu:

- dočasné úložiště – úložiště pro elektronické dokumenty pro jejich přijetí, ověření, atd.
- dlouhodobé úložiště – pro dlouhodobé uchování elektronických dokumentů
- přístupové úložiště – pro uložení elektronických dokumentů, ke kterým byl požadován přístup uživatelů archivu

[51] Viz např. Jim Linden, Sean Martin, Richard Masters, and Roderic Parker (The British Library): „The large-scale archival storage of digital objects“, DPC Technology Watch Series Report 04-03, February 2005

[52] Reich, V & Rosenthal, D.: Lockss: A permanent publishing and web access system. D-Lib Magazine, 7, (6), (2001)

Je zřejmé, že požadavky pro tyto jednotlivé typy využití se budou lišit, proto také architektura je musí zohlednit pomocí různých úložných systémů s různými typy medií, ale zároveň je nutné, aby tvořili jeden integrovaný systém.
Podobné zadání řeší systémy HSM (Hierarchical Storage Management), které kombinují několik typů úložných zařízení (např. diskové pole s páskovými knihovnami) do jednoho logického systému, který efektivně a ekonomicky využívá jednotlivá zařízení (např. dlouho nepoužívané soubory se přesunou z diskového pole na pásky). HSM zajišťuje rozmístění a přesun dat mezi jednotlivými zařízeními automaticky, na základě předem stanovených pravidel. HSM může být jednou z možností jak zefektivnit a zabezpečit datové úložiště archivu[53].

2.3.4.2 Výběr vhodného typu medií

Elektronická média mají z hlediska archivace tu vlastnost, že doba jejich uchování je omezena. Je to dáno jednak fyzickými vlastnostmi média (kde současná média mají životnost v řádu několika desítek let), ale také vývojem technologií, kdy zařízení a programy pro čtení z daných medií zastarají, přestanou být dostupná a tím pádem nejsou použitelná ani média samotná. Právě tento druhý faktor je určující – z hlediska vývoje technologie bude nutné (a i ekonomické) migrovat spíše data z jedné technologie na jinou, než prodlužovat životnost jednoho typu medií (i když i to je důležité jak je popsáno v následující podkapitole).

I když je tedy jisté, že média se budou měnit během archivace, jejich vhodný výběr je důležitý a má dopad jak na funkční tak i ekonomickou stránku archivu.
Mezi hlavní kriteria pro výběr medií patří:

- životnost
- kapacita
- rychlost čtení, zápisu a vyhledání
- odolnost
- cena (pořizovací cena a hlavně celková cena vlastnictví v přepočtu na jednotku dat a času)
- stárnutí technologie

Pro uložení elektronických dokumentů přicházejí v úvahu tyto hlavní média:

- pevné disky
- optické disky
- magnetické pásky

Jiné typy médií jako magneto-optické disky nebo paměti flash nejsou pro toto použití rozšířeny.

Následující tabulka podává krátké srovnání jednotlivých medií (současný stav):

	Pevné disky	Optické disky	Magnetické pásky
životnost	nejnižší (5-15 let)	nejvyšší (~ 50 let)	střední(~ 30 let)

[53] Viz např. zkušenosti s HSM popsané v Ronald Jantz, Michael J. Giarlo: „Digital Preservation - Architecture and Technology for Trusted Digital Repositories“, D-Lib Magazine, June 2005, Volume 11 Number 6

kapacita	až 800 GB	až 50 GB	až 400GB
rychlost	nejlepší	nejpomalejší, ale rychlé vyhledávání	prostřední, ale velmi pomalé vyhledávání
cena	nejdražší	prostřední	nejlevnější

Pevné disky

Pevné disky jsou nejrychleji přístupným mediem jak pro čtení tak i zápis. Z hlediska životnosti jsou vhodné spíše pro střednědobé nebo krátkodobé uložení, i když tento přístup se částečně mění s rozvojem diskových polí.
Právě disková pole s technologií RAID (Redundant Array of Independent Disks) jsou často využita pro uložení dat v archivech. Technologie RAID zajistí vyšší spolehlivost a životnost, než mohou zajistit samostatné disky.
Z hlediska dlouhodobého ukládání je vhodná např. RAID 6, která umožní zvládnout výpadek až dvou disků v jedné skupině.
K širšímu využití diskových polí také přispívá rozvoj technologie pevných disků SATA, kdy vysokokapacitní pevné disky jsou dostupné za zajímavé ceny.
Právě kombinace SATA disků a vhodné RAID technologie, může nabídnout vhodné úložiště, kdy cena bude bližší jiným řešením, dříve výrazně levnějším.
Zajímavá je zde i vysoká integrace kapacity – kdy např. standardní 19“ počítačový rack s diskovými poli SATA může nabídnou kapacitu okolo 80TB.

Současný vývoj na tomto poli přichází s další technologií - MAID (Massive Array of Idle Disks), která navazuje na technologie RAID a která řeší jednu z nevýhod diskových polí pro dlouhodobé uchování dat – a to že všechny disky v poli musí být zapojeny a aktivní, i když nejsou delší dobu využity. MAID technologie právě umožňuje vypnout nepoužívané disky a tím snížit především spotřebu elektrické energie a také částečně zlepšit životnost disků.

Optické disky

Přestože optická média jsou poměrně velmi rozšířená jako nosič pro šíření audio a video souborů, jako řešení pro masové ukládání dat jsou používány méně často a spíše v případech, kdy je potřeba využít některou jejich specifickou vlastnost (jako např. WORM – nepřepisovatelný zápis viz detailní popis WORM technologií dále; jedině optická média nabízí WORM jako vlastnost medií). Důvodem je menší kapacita medií, takže pro větší objemy dat se stává správa medií obtížnější (zatímco kapacita páskové knihovny může jít až do pentabytů, optické knihovny se pohybují do desítek terabytů).
Optická média zahrnují známé technologie CD a DVD a nové technologie pro vysokokapacitní optické disky UDO a PDD (které jsou v současné době využívány hlavně pro ukládání dat) a Blue Ray a HD DVD (které se zatím zaměřují více jako nosiče pro digitální video). Následující tabulka uvádí pro srovnání kapacitu těchto různých medií:

Medium	CD	DVD	PDD	UDO	HD DVD	Blue Ray
Kapacita	700MB	9GB	23GB	30GB	30GB	50GB

Magnetické pásky

Magnetické pásky patří mezi nejrozšířenější řešení pro zálohování dat a jsou také často využívána jako medium pro archivaci. I když jejich životnost je kratší, než u optických medii, nabízí vhodnou ekonomickou a technickou alternativu hlavně díky rozšíření a podpoře výrobců.
V současné době jsou dostupné dvě hlavní technologie pro magnetické pásky SDLT a LTO (Ultrium), kdy technologie LTO začínají výrazně převládat.
Kromě těchto hlavních směrů existují technologie (jako např. nabízí Sun-StorageTek), které se zaměřují na specifické aspekty – vysoká výkonnost a dostupnost, šifrování obsahu na úrovni mechaniky atd.
Následující tabulka udává porovnání hlavních generací LTO a SDLT pásek:

Technologie	LTO-1	LTO-2	LTO-3	LTO-4*	SDLT 220	SDLT 320	SDLT 600	DLT S4*
Kapacita (GB)	100	200	400	800	110	160	300	800
Rychlost max (MB/s)	20	40	80	120	10	16	36	60

* příští generace, v součastné době ve vývoji.

Kromě vlastního média je samozřejmě také důležitá mechanika, která na pásku zapisuje a nebo z ní čte data. Moderní mechaniky, jak pro LTO, SDLT, tak pro obdobné privátní technologie jsou komplexní zařízení, které obsahují mikroprocesory s rozsáhlým firmware, který přímo v mechanice zajišťuje funkce jako komprimaci dat a opravu dat (kterou jsou již vybaveny prakticky všechny mechaniky), nebo šifrování a WORM funkcionalitu (které jsou dostupné v nejnovějších nebo připravovaných technologiích).

Zařízení, která umožňují práci s větším množstvím pásek najednou ze nazývají páskové knihovny. Tyto obsahují několik stovek až tisíců pozic pro uložení pásek, až několik desítek páskových mechanik a robotický manipulátor, který přemísťuje pásky z úložných pozic do mechanik a zpět. Za využití páskové knihovny a vhodného software může mít informační systém přístupné zcela automaticky až pentabyty dat.

Další zajímavou technologií blízkou jak magnetickým páskám tak diskovým polím jsou VTL (Virtual Tape Libraries), kdy funkce páskové knihovny je emulována diskovým polem. VTL pak umožňuje rychlejší a snadnější manipulaci s daty při zachování standardního rozhraní, a může sloužit jako dočasné úložiště pro zefektivnění práce s páskovou knihovnou.

WORM
WORM technologie (Write Once Read Many), je důležitou technologií pro archivování elektronických dokumentů. Tato technologie umožňuje na daném médium zapsat pouze jednou a zapsaná data pak nelze změnit ani smazat. WORM

tedy zvýší důvěryhodnost uložených dat, neboť zaručí, že data nebyla změněna od svého uložení. [54]
Původně byla WORM média pouze optická, kde fyzické vlastnosti média zaručují, že na ně lze zapsat pouze jednou. V současné době existují i WORM magnetická média – pásky a i pevné disky, kde je nepřepisovatelnost zaručena na úrovni firmwaru.

2.3.4.3 Proces kontroly a údržby medií během uchování

Všem typům medií musí být věnována během uložení vhodná péče již od nákupu, vstupní kontroly, inicializace přes uložení dokumentů k uchovávání až po kontroly stavu medií. Některé níže uvedené kroky se používají především pro vyměnitelná média.

Kroky, které mají být prováděny, před použitím medií:

- Média pro archivaci by měla být nakupována pouze od osvědčených výrobců
- Po nákupu by měly být média ověřena (náhodný výběr z jednotlivých sérií), zda splňují požadované technické parametry (tak lze odhalit např. výrobní chybu v dané sérii)
- Pro daná média stanovit maximální dobu životnosti na základě dostupných informací od výrobců, standardů, výsledků nezávislých a vlastních testů.

Kroky, které mají být prováděny při použití medií:

- Medium má být označeno jednoznačným sériovým číslem, které navazuje na sériové číslo předchozího média (sériové číslo může být také součástí inicializačních dat – viz níže); sériové číslo by mělo být uvedeno i na obalu média
- Inicializace média – pokud to dané medium umožňuje, před jeho použitím by mělo být inicializováno následujícími daty:
    - identifikace zařízení, které provedlo inicializaci – HW i program
    - identifikace organizace, osoby
    - datum a čas
    - sériové číslo a sériové číslo předchozího média
- Po uložení dokumentu, vždy provést kontrolu, zda data byla zapsána na medium přesně, bez chyby
- Jako dodatečná metadata může být s dokumenty uloženo i jednoznačné výrobní číslo mechaniky, na které probíhal zápis
- Uzavření média - po uložení všech dokumentů na médium je vhodné provést uzavření média, kdy se na médium zapíše:
    - identifikace zařízení, které provedlo uzavření – HW i program
    - identifikace organizace, osoby
    - datum a čas
    - sériové číslo následujícího média
- Auditní záznamy (pořizované archivním systémem) by měly obsahovat údaje o inicializaci média, uložení dokumentů a uzavření média (pokud média

[54] Např. ISO-TR 15801 doporučuje ukládat na WORM média auditní záznamy o elektronicky uchovávaných dokumentech. ISO-TR 18509 považuje ukládání na WORM za výrazný prvek zabezpečení uložených záznamů (jako kompromisní řešení umožňuje ukládat na WORM pouze otisky záznamů), v případě uložení záznamů u třetí strany požaduje použití WORM medií.

obsahuji jednoznačné výrobní číslo, mělo by toto být referováno v auditních záznamech).

Kroky, které mají být prováděny při uchovávání médií:

- Média musí být uložena ve vhodném prostředí, které splňuje podmínky pro jejich dlouhodobé uchování (podmínky jsou uvedeny výrobcem a také některými normami), jedná se především o:
  - teplotu
  - vlhkost
  - prašnost
  - světelné podmínky
  - magnetické pole
- Pro uchovávaná média provádět pravidelnou kontrolu kvality médií, kdy pomocí vhodného testovacího zařízení, jsou měřeny údaje o kvalitě čtení dat z média (frekvence výskytu opravitelných chyb, úroveň signálu aj.) a porovnávány s hodnotami uváděnými výrobci nebo doporučenými standardy[55]
- Pokud hodnoty naměřené při kontrole média přesáhnou povolené hodnoty je nutné přenést data okamžitě na nové médium a případně také upravit informace o sekvenci medií tak, aby bylo zřejmé, kam nahrazené medium patří
- Pokud je na mediu nalezena neopravitelná chyba použít repliky na jiných mediích, pomocí kterých budou nahrazeny nečitelné dokumenty a vytvořena identická kopie poškozeného média
- Taktéž automaticky nahrazovat média, která přesáhnout stanovenou dobu životnosti
- O kontrolách medií a jejich nahrazeních je nutné vést auditní záznamy

2.3.4.4 Zestárnutí technologie úložiště dat přechod na nový systém

Jak již bylo uvedeno výše, kromě vlastního stárnutí medií, dochází k nahrazování starých technologií novými a po určité době je pak obtížné zajišťovat technické prostředky, které umožňují čtení ze starých médií.

Tento jev si můžeme ukázat například na technologii magnetických pásek LTO. Nová generace technologie se uvádí na trh každé 2-3 roky (v současné době je běžně využívána LTO-3 a LTO-4 bude brzy dostupná).
Pro zpětnou kompatibilitu LTO generací platí následující pravidla:

a) daná generace umí číst pásky alespoň o dvě generace zpátky
b) daná generace umí zapisovat na pásky o generaci starší (starším formátem)

Pokud se podíváme na trh, vidíme, že běžně jsou dostupné produkty pro 2 generace pásek – nejnovější a předchozí. Z výše uvedeného vyplývá, že pokud máme implementováno řešení na technologii o 4 generace starší než je současná, bude pro nás obtížné je udržovat. Pokud jsme začali s technologií v dané době aktuální, tato situace nastane za cca 8-12 let. Také pokud porovnáváme vývoj u jiných technologií

[55] Pro optická média je např. dostupná norma ISO 12142, Electronic imaging – Method of monitoring and reporting errors in data stored on digital optical disks

uložení dat ukazuje se, že doba okolo 10 let odpovídá efektivní životnosti těchto technologií.

Pokud porovnáme dobu životnosti médií (kterou výrobci uvádí cca 30 let pro pásky a až 50 let pro optické disky), je zřejmé, že bude nutné dříve vyměnit zařízení pro ukládání dat a provést migraci starých medií na novou technologii než bude nutné obnovit většinu starých médií z důvodů vypršení jejich životnosti.

Přechod na nový typ úložné technologie (a s tím nová zařízení a nová média) může být komplexní, rozsáhlý i finančně náročný projekt, proto jej archív musí plánovat s dostatečným předstihem.

S přechody mezi různými technologiemi také úzce souvisí architektura úložiště dat. Tato musí počítat s budoucími přechody na jiné technologie, musí umožnit v rámci systému vybudovat úložiště založené na nové technologii a plynulý přechod starých částí systému na nové při plném provozu a zajištění plné autentičnosti a integrity přenesených dokumentů.

Mezi hlavní oblasti, na které by se měl projekt přechodu zaměřit, jsou:

- podrobná riziková analýza
- omezit přechod pouze na jednu repliku archivních dokumentů závislou na dané technologii, zachovat ostatní kopie nedotčené
- nový systém musí mít rozhraní, která budou navazovat na architekturu archivu a umožní jeho integraci do celkového řešení
- dostatečné ověření a otestování nového systému
- bezchybné přenesení všech dat na nová média a kontrola přesného přepisu dokumentů
- zachování referencí na média, aktualizování referencí (nová média budou mít typicky několikanásobně větší kapacitu, je potřeba provést jednoznačné přiřazení starých referencí novým, aktualizovat sekvenci medií a zaručit, že všechny reference na média budou platné)
- o plánování a jednotlivých krocích přechodu na novou technologií včetně přenesení jednotlivých dokumentů na nová média je nutné vést detailní auditní záznamy

2.3.5 Zabezpečení dat, metadat a komunikačních protokolů během životního cyklu elektronického dokumentu

Potřeba dlouhodobě uchovávat digitální data jako průkazný materiál vzniká v rámci celé řady praktických situací. Často je proto zapotřebí, aby takováto archivace splňovala určité specifické podmínky. Například důležitým může být hledisko bezpečnosti a trvalosti dat, může (při důkazním řízení) být nutné prokázat, že tato data resp. přidružená metadata existovala v určitém čase v minulosti a že nebyla od té doby změněna. Pokud za tímto účelem byly použity postupy jako digitální podpis nebo časové razítko a konkrétní kryptografické algoritmy, je nezbytné prokázat, že podpis či časové razítko existovaly předtím než se použité kryptografické algoritmy staly slabými (např. z hlediska délky použitého klíče nebo nalezení kolizí u použité hašovací funkce apod.) či dříve než příslušné certifikáty vypršely či byly odvolány.

Při dlouhodobé archivaci v rámci životního cyklu ED provádět takové aktivity, které umožní uchovat nepopiratelnost existence a nenarušenost dat a stejně tak zajistit i jejich požadovanou dostupnost. Za tímto účelem je nutné použít celou řadu technických a operačních prostředků, které přesahují kryptografický rámec - média pro ukládání dat, plány pro případ živelných pohrom, změny technologií pro zpracování dat, legislativní požadavky atd.

V tomto odstavci se podíváme na některé navržené protokoly nebo postupy, které za použití standardizovaných kryptografických protokolů a algoritmů, umožní splnění těchto speciálních požadavků jakož i klasických bezpečnostních požadavků, při ukládání dat a jim přidružených metadat po celou dobu životního cyklu ED v archívu.

Navržené protokoly se snaží zejména vytvořit všechny nezbytné předpoklady pro dlouhodobou „funkčnost“ (hledisko ověřování) technologií digitálního podpisu a časového razítka, se kterými se počítá jako se základními stavebními kameny.

Scénáře aplikací

V řadě praktických situací se lze setkat s problémem jak uložit bezpečným způsobem elektronické dokumenty na nějakou často neurčitou dobu. Jedná se např. o digitální smlouvy, daňová přiznání, faktury nebo elektronické rodné listy. Zde jde nejen o to je uchovávat, ale také je důležité prokázat, že tyto dokumenty existovaly v určitém čase v minulosti a že od té doby nebyly pozměněny.

V průběhu času však může docházet k tomu, že průkazní hodnota elektronických podpisů či časových razítek klesá či dokonce mizí a to z řady příčin:

- již není dostupná příslušná informace o zneplatnění (přestala pracovat odpovídající CA či protokol OCSP);
- nejsou dostupné certifikáty, které jsou zapotřebí pro verifikaci elektronického podpisu;
- certifikát asociovaný s elektronickým podpisem vypršel či byl odvolán;
- vývoj kryptografických či výpočetních technik již pokročil tak, že lze vnutit dokumenty či podpisy nebo spočítat utajované soukromé klíče.

Abychom se těmto problémům vyhnuli, je vhodné navrhnout protokol jak uchovávat spolu s dokumenty i příslušné průkazní záznamy (certifikáty, CRL, odpovědi OCSP a časové značky), periodicky obnovovat potřebné záznamy a přidávat další nezbytné informace - například nové časové značky používající silnější algoritmus.

2.3.5.1 DVCS (Data Validation and Certification Server Protocols)

V souvislosti s výběrem vhodného ověřovacího protokolu pro archivní služby bývá diskutován a používán protokol DVCS.

Tento protokol vznikl v rámci pracovní skupiny PKIX. Popsán je v RFC 3029, které bylo zařazeno do kategorie experimentálních a není tedy běžnou de-facto internetovou normou. Na jeho vzniku se podíleli pracovníci firem Entrust (C.Adams a R.Zuccherato), Baltimore (M. Zolotarev) a EdelWeb SA (P. Sylvester).

Dokument RFC 3029 popisuje vlastnosti severu DVC (Data Validation and Certification Server) a protokoly používané pro komunikaci s tímto serverem. DVC server je chápán jako důvěryhodná třetí strana, která může být využívána pro vytváření služeb, kde je vyžadována nepopiratelnost (non-repudiation services). Server vytváří určitá svědectví - potvrzení (ve vztahu k platnosti elektronicky podepsaných dokumentů, certifikátů veřejných klíčů, resp. existencí určitých dat), tato potvrzení nazývá ověřovacími certifikáty (Data Validation Certificates, dále jen DV certifikáty). Tyto certifikáty pak lze použít při konstrukci příslušných důkazů (platnost a nepopiratelnost elektronických podpisů, ověření, že příslušný uživatel vlastní určitá data, ověření platnosti a revokačního statutu certifikátu veřejného klíče atd.). U uložených dat přítomnost DV certifikátu prokazuje, že příslušný digitálně podepsaný dokument či certifikát veřejného klíče byly platné v čase, který je obsažen v DV certifikátu.

V materiálu jsou definovány 4 typy ověřovacích a certifikačních služeb:

- certifikace vlastnictví (possession) dat (cpd);
- certifikace tvrzení o vlastnictví dat (ccpd);
- ověření platnosti digitálně podepsaného dokumentu (vsd);
- ověření platnosti certifikátu veřejného klíče (vkpc).

Transakce s DVC serverem začíná přípravou klientské žádosti, tato žádost vždy obsahuje data, jejichž platnost, vlastnictví či správnost je ověřována. Je zvolen vhodný transportní mechanismus (umožňující zajistit důvěrnost transakce, autentizaci DVC serveru např. pomocí TLS či CMS nebo pomocí šifrování S/MIME).

Server DVC po obdržení žádosti ověří její platnost a provede odpovídající ověřovací postupy. Následně (v případě shody) provede vygenerování DV certifikátu a pošle odpověď, která obsahuje tento certifikát.

Zbývající část RFC obsahuje konkrétní syntaxi ASN 1 výše popsaných objektů a zpráv.

V současnosti je však tento protokol podroben poměrně důrazné kritice. Faktickou příčinou kritiky je skutečnost, že problematika ověřování platnosti certifikátů (včetně souvisejících otázek) za poslední dva roky značně pokročila. To se však nedá říci o DVCS problematice, která ustrnula na stavu z počátku roku 2001.

[1] Adams, C.; Sylvester, P.; Zolotarev, M.; Zuccherat, R.: Data Validation and Certification Server Protocol, RFC 3029, February 2001

2.3.5.2 Protokol TAP (Trusted Archive Protocol)

Kritika předchozího protokolu DVCS vedla k návrhu nového protokolu Trusted Archive Protocol (draft-ietf-pkix-tap-00.txt). Autory protokolu jsou C. Wallace (CygnaCom Solution – dceřiná společnost Entrustu) a S.Chokhani (Orion Security – konzultační firma). Dokument vznikl v rámci významného projektu Long-term Archive and Notary Services (LTANS).

Dokument popisuje službu důvěryhodné archivní autority (TAA – Trusted Archive Authority), která má sloužit jako podpora dlouhodobé nepopiratelnosti pomocí bezpečného uložení (kryptograficky obnovované) informace. Tato služba (TAA) zajišťuje dlouhodobé uchovávání dat pomocí obnovy časových značek. Pro tento účel je definován důvěryhodný archivní protokol (TAP - trusted archive protocol), který umožňuje interakci s TAA.

Entity, které přispívají do archivu se nazývají přispěvatelé; entity, které naopak vyžadují přístup k datům, či vyžadují odstranění určitých dat, se nazývají žadatelé.

Objekty archivačního procesu jsou v rámci dokumentu členěny následovně:

- archivovaná data - data, která jsou přispěvatelem zasílána TAA;
- archivní známka – objekt generovaný TAA po obdržení dat od přispěvatele a akceptaci jejich archivace. Je zasílána zpět k přispěvateli a lze ji použít k žádosti o vyhledání či odstranění archivovaných dat a asociovaných informací kryptografického charakteru. Známky obsahují: přispěvatelovo DN, časovou značku, datum a čas, kdy příspěvek TAA obdržela a (nepovinně) vyhledávací informace. Přispěvatel musí ověřit obsah této známky (pro ověření správnosti složení uložených informací v TAA);
- archivní záznam, obsahuje kryptografickou obnovu historie, kterou provádí TAA. Prvotní archivní záznam – to je časová značka, která byla spočtena pro data získaná od přispěvatele. Formát časové značky je definován podle RFC 3161. Při každé obnově nově získaný archivní záznam obsahuje předešlou verzi tohoto archivního záznamu a novou časovou značku. Při ověření archivního záznamu je toto ukončeno v momentu, kdy je ověřena původní časová značka.
- archivní soubor (package), je složený objekt, který minimálně obsahuje archivní známku, archivní záznam a archivovaná data. Může obsahovat další kryptografické informace.

TAA potenciálně může archivovat data v libovolném formátu, je však také možné, že na typ archivovaných dat bude kladeno nějaké omezení. Data přispěvatelů mohou obsahovat veškerou potřebnou kryptografickou informaci, mohou obsahovat pouze její část, resp. nemusí obsahovat žádnou.

V rámci protokolu TAP vystupují následující čtyři typy entit: TAA, TSA, klient-přispěvatel a klient-žadatel.

TAA musí zabezpečovat následující služby:

- uchovávání archivovaných dat;
- generování archivních známek (včetně vyžádání časové značky pro archivovaná data);
- periodickou obnovu archivních záznamů;
- uchovávání svěřených kryptografických informací pro verifikaci archivního záznamu (kořenový certifikát poskytovatele, použité certifikáty, CRL, odpovědi protokolu OCSP, certifikáty serveru OCSP atd.);
- přijetí archivního souboru a jeho odstranění.

TAA dále může provádět další nepovinné služby jako např.:

- uchovávání historických kořenových certifikátů důvěryhodného poskytovatele;
- sběr a ověřování informací ve vztahu k PKI;
- ověřování kryptografických zpráv.

Návrh protokolu se opírá (stejně jako RFC 3126 – formáty dlouhodobě platných elektronických podpisů) o syntaxi CMS (RFC 3369) a protokol pro časové značky (RFC 3161).
Definuje

- protokol pro přenos dat mezi TAA a klienty;
- objekty, které lze používat k archivaci a uchovávání libovolné kryptografické služby, jako je digitální podpis a k archivaci libovolných nekryptografických dat.

Protokol TAP používá přístup opírající se o obnovu časových značek a tak značně snižuje nároky na stupeň důvěry k TAA vzhledem k integritě archivovaných dat. Jinými slovy, případné modifikace dat v archivních záznamech TAA mohou být detekovány.

TAA se nemusí zabývat archivovanými daty (z hlediska obsahu) a lze ji využít k archivaci jak kryptografických tak i nekryptografických dat. Kryptografická data mohou být buď podepsána či šifrována anebo jsou současně šifrována a podepsána.

Pro podporu dlouhodobého uchovávání elektronických podpisů může přispěvatelem zasílaný soubor obsahovat všechny certifikáty, revokační informace (odpovědi CRL a OCSP) a také kořenový certifikát poskytovatele, aby byla usnadněna následná verifikace v libovolném budoucím čase a to bez potřeby obracet se k službám či skladům nebo k jiným zdrojům informací o certifikátech a revokacích.

Pokud pak žadatel používá jiný důvěryhodný zdroj k ověření podpisu a časových značek, pak není nutné opírat důvěru v integritu dat o důvěru v TAA. Autorita časových značek však musí být důvěryhodnou institucí v každém případě.

Požadavky přispěvatelů a žadatelů mohou být buď podepsány nebo mohou být neautentizovány, či autentizovány jinými prostředky (např. klientská autentizace pomocí SSL/TLS). Požadavky na odstranění části archivu musí být autentizovány. Zprávy TAA jsou vždy podepsány (CMS SignedData – RFC 3369). Odpovědi TAA musí být vždy podepsány a nesmí obsahovat jakýkoliv jiný podpis (kromě podpisu samotné TAA). V odpovědích musí být obsažen certifikát TAA serveru. Mohou zde být i další certifikáty a případná CRL.

Materiál definuje formáty žádostí (přispěvatelovy, žadatelovy), formáty odpovědí a jejich konkrétní ASN.1 syntaxi.

Podpisy všech odpovědí TAA musí být ověřovány, přitom postupy se mírně liší v závislosti na druhu prováděné transakce. V dokumentu je popsáno využití http jako přenosového protokolu. Žádný konkrétní typ přenosového protokolu však není stanoven jako povinný.
V páté kapitole se autoři zabývají některými možnými přístupy ke kontrole archivu. Kontroly obvykle probíhají na základě klientského požadavku (jsou součástí žádosti).

Vzhledem k tomu, že se vlastně jedná teprve o návrh, jsou zde některé další momenty práce TAA zatím spíše jen naznačeny. To se týká například problematiky autorizace (kdo a jak má právo používat služeb TAA), nezbytných „bezpečnostních" vlastností TAA – potřebných k tomu, aby TAA mohla fungovat jako důvěryhodná strana (zpracování dokumentace, využívání kontrolních prostředků, fyzické prostředky vhodné pro archivaci, atd.).

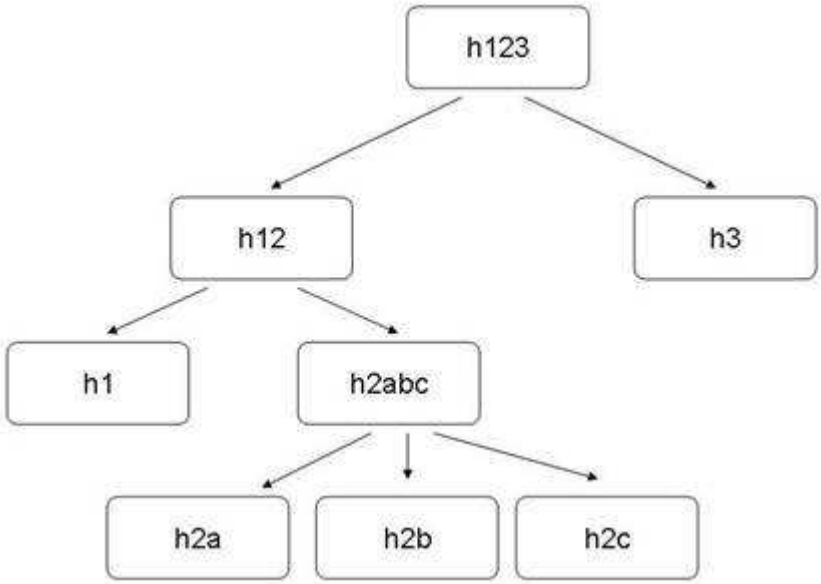
[1] CMS, RFC 3369, August 2002

[2] Pinkas, D., Ross, J., and N. Pope, Electronic Signature Formats for Long Term Electronic Signatures, RFC 3126, September 2001.

[3] Adams, C., Cain, P., Pinkas, D. and R. Zuccherato, "Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP)", RFC 3161 , August 2001.

2.3.5.3 ERS (Evidence Record Syntax)

Existuje více scénářů jak v archivu řešit otázku existence a integrity dat v případě, kdy je nutné provést změny v používaných kryptografických algoritmech (náhrada „slabých“ algoritmů) resp. v případě, že je nutné částečně porušit celkovou integritu archivu (mazání některých nepotřebných dat). V poslední době se do popředí dostává metoda zřetězených otisků. K této metodě existuje řada teoretických studií a prací. Z hlediska praxe je významným krokem k rutinnímu nasazení standardizační proces v rámci LTANS (Long-term Archive and Notary Services). Výsledkem těchto aktivit je návrh, který byl publikován začátkem ledna tohoto roku (2007) s názvem Evidence Record Syntax (ERS). Precizuje vhodná doporučení jak vytvářet „stromy“ zřetězených otisků jednotlivých objektů a zejména jak provádět následnou redukci, pokud je vyžadována. Práce s otisky je doprovázena vhodnou podporu časových razítek.



Na obrázku je příklad zřetězení otisků dat některých objektů. Postup dokumentuje, že k prokázaní integrity dat souborů a,b,c stačí úschova otisku obecnějších objektů 12 a

3. Toto není nic „nového“, neboť je zřejmé, že pro důkaz integrity celého archívu (a tedy všech uložených objektů) stačí uchovávat jen jeden otisk – otisk celého archívu. Přínos popsané metody je tedy v něčem jiném a to v tom, že popisuje možnost, jak přejít na související strukturu, kde se otisky vytváří jinými algoritmy, a v tom, že existuje sofistikovaný postup, jak redukovat (v případě potřeby) některé větve stromu s tím, že se neohrozí průkaznost zachování integrity ostatních objektů.

[1] http://tools.ietf.org/wg/ltans/draft-ietf-ltans-ers/draft-ietf-ltans-ers-09.txt , January 04, 2007

2.3.5.4 LTAP (Long-term Archive Protocol)

Protokol LTAP (Long-term Archive Protocol) řeší obecné požadavky na službu dlouhodobé archivace. Je popsán v materiálech pracovní skupiny LTANS. V současné době byla přijata již čtvrtá verze tohoto dokumentu (26.10.2006). Jedná se však stále jen o návrh (s platností do dubna 2007). Dle jejich autorů (A. Jerman-Blazic, P. Sylvester, Wallace, C.) se předpokládá jeho rozšíření a úprava některých částí materiálu (speciálně části odvolávající se na nově připravované řešení zachování průkaznosti, viz např. ERS) .

Protokol se zabývá zejména „konzervací” a „dematrializací” přijímaných dokumentů a to v protokolu klient – server.

[1] http://www.ietf.org/internet-drafts/draft-ietf-ltans-ltap-03.txt 23.10.2006

2.3.5.5 Long-term Archive Service Requirements (Požadavky na službu dlouhodobé archivace)

Cílem dokumentu Požadavky na službu dlouhodobé archivace (Long-term Archive Service Requirements) jakožto jednoho z úvodních materiálů pracovní skupiny LTANS je specifikovat technické požadavky, které souvisí s poskytováním služby dlouhodobé archivace podporující možnosti ubezpečovat se (a prokazovat) existenci a nenarušenost dat, speciálně digitálně podepsaných dat.

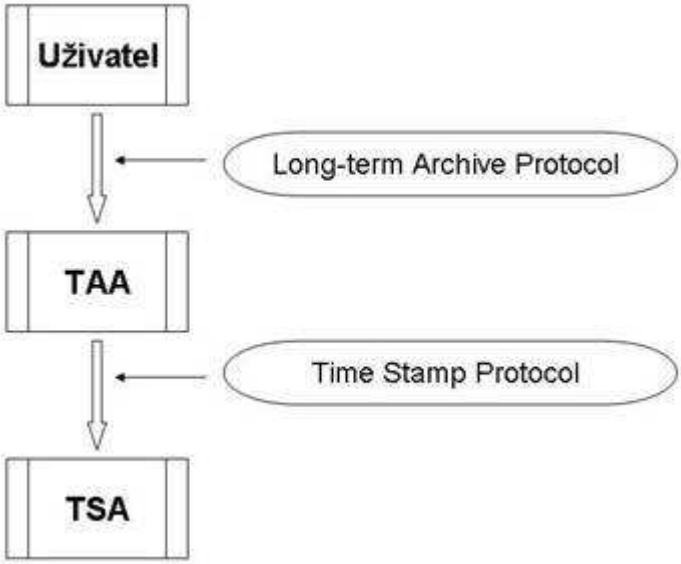
Služba dlouhodobé archivace je navržena tak, aby řešila podstatné části problémů zmíněných v úvodu k této části. Archivované datové objekty budou ukládány na dlouhá resp. nedefinovaná časová období a bude použita celá řada prostředků, které budou garantovat dostupnost těchto dat (plány obnovy, redundatní ukládání, havarijní plánování).

Služba poskytuje materiál nezbytný k prokazování existence a nenarušenosti dat a to jak ve vztahu k uživatelům, tak i ve vztahu k jurisdikci. Poskytuje prostředky pro uchování průkazních záznamů pro podepsané archivované datové objekty. Neřeší však všechny myslitelné problémy související s dlouhodobou verifikací elektronických podpisů - např. neposkytuje prostředky pro ověření podpisů, které jsou přímo součástí archivovaných datových objektů. Toto provádí ověřovací služby v rámci PKI jako SCVP či DVCS. Neposkytuje také prostředky, jak zahrnout ověřovací

data do datových objektů. To zase by mělo být prováděno dle jiných specifikací (např. dle RFC 3029 a s ohledem na formát dokumentu).

Naopak služba dlouhodobé archivace poskytuje prostředky pro nepopiratelnost v dlouhém časovém období prostřednictvím periodického vytváření časových razítek.

Základní funkce služby dlouhodobé archivace jsou realizovány v několika instancích:



Uživatel převádí datové objekty, které by měly být archivovány důvěryhodnou archivní autoritou (Trusted Archive Authority - TAA) a používá k tomu aplikace dle svého výběru. Prostřednictvím protokolu služby dlouhodobé archivace a specifikací formátu dat balíku archivu pak uživatel může požadovat uložené datové objekty a asociované průkazní záznamy. TAA uloží dokumenty a získá k nim nezbytná ověřovací data (speciálně časová razítka prostřednictvím protokolu pro časová razítka – RFC 3161, resp. prostřednictvím jiných protokolů jako SCVP získá další ověřovací data). TAA může poskytovat služby TSA (autority časových značek), je zde však nezbytné určité rozlišení. Jako TAA může sloužit server uvnitř počítačové sítě organizace, který využívá lokální archivní servery nebo i vnější služby dosažitelné prostřednictvím Internetu.

Funkční a kvalitativní požadavky na službu dlouhodobé archivace

Služba dlouhodobé archivace musí poskytovat následující základní funkce:

- přijímat datové objekty či skupiny datových objektů pro jejich uchovávání;
- ukládat převzaté datové objekty pro dané časové období;
- vytvářet, ukládat a provozovat průkazní záznamy (např. prostřednictvím periodického získávání časových razítek) pro datové objekty, které byly převzaty pro uchování;
- sbírat a ukládat další ověřovací data nezbytná pro ověření průkazních záznamů;

- poskytovat balíky archivu obsahující archivovaná data, průkazní záznamy či oboje;
- poskytovat služby podle politiky dlouhodobé archivace;
- být schopna poskytovat archivní balíky i v případě, že se technologie pro ukládání či technologie pro zpracování změnily během života archivovaného datového objektu;
- být schopna poskytovat informace, že datový objekt existoval v určité době jako alternativu v situacích, kdy uživatel není schopen interpretovat průkazní záznamy;
- fungovat podle archivační politiky, která jako minimum stanoví kvalitu časových razítek a podmínky pro jejich obnovu, atd.

Služba dlouhodobé archivace musí být schopna efektivní činnosti i pro obrovská množství archivovaných datových objektů. Je proto třeba minimalizovat příslušné objemy činností (počty časových razítek, přístupy k archivovaných datovým objektům atd.).

Požadavky na strukturu archivovaných dat

Datová struktura balíku archivu má obsahovat archivovaný datový objekt a průkazní záznam.

Struktura průkazního záznamu by měla mít následující vlastnosti:
- musí být umožněno zahrnutí všech časových značek, které jsou nezbytné pro ověření existence archivovaného datového objektu;
- struktura časového razítka by měla umožnit efektivní poskytnutí důkazů mnoha archivovaných datových objektů;
- mělo by být umožněno poskytnutí důkazů pro skupiny archivovaných datových objektů;
- -pokud jsou předloženy k archivaci skupiny datových objektů, musí důkaz nepopiratelnosti být dostupný i odděleně pro každý archivovaný datový objekt;
- odstranění některých archivovaných datových objektů nesmí vést k rizikům při důkazech pro jiná archivovaná data;
- musí být možné vytvořit časová razítka bez nutnosti přístupu k samotným archivovaným datovým objektům. Takováto nezbytnost přístupu k archivovaným datovým objektům může vzniknout pouze v případě, že jsou narušeny bezpečnostní vlastnosti použitého hašovacího algoritmu;
- všechny v čase použité hašovací algoritmy musí být identifikovány (v jednom místě) tak, aby bylo umožněno jednorázové ověřování;
- další požadavky se týkají vytváření balíků obsahujících průkazní záznamy, zašifrovaných archivních datových objektů resp. zařazení dalších informací.

Požadavky vzhledem k protokolu pro interakci se službou dlouhodobé archivace

Tento odstavec dokumentu zvažuje nároky, které by měl splňovat protokol pro komunikaci se službou dlouhodobé archivace. Protokol musí logicky zabezpečit základní funkce služby jako jsou předkládání datových objektů k archivaci, přístup k balíkům archivu, odstraňování dat resp. průkazních záznamů z archivu. Musí také

vyhovět některým bezpečnostním nárokům a umožnit i práci s průkazními záznamy, které vytvořila jiná TAA.

Literatura:
[1] webovská stránka ltans: http://ltans.edelweb.fr/
[2] Long-term Archive Service Requirements
http://ltans.edelweb.fr/draft-ietf-ltans-reqs-05.txt , říjen 2005
[3] J.Pinkava: Archivace elektronických dokumentů, Crypto-World 04/2002

2.3.6 Rámcová architektura systému archivu

Archivační systém je komplexní řešení, kde je integrována řada hardwarových a softwarových komponent. Prakticky neexistují hotová řešení, každý systém je řešen jako individuální projekt, kde se integrují dostupné komerční a nekomerční systémy s vyvinutými softwarovými moduly tak, aby vzniklo řešení, které unikátně pokrývá požadavky daného archivu. Je tedy těžké hovořit o nějaké vzorové architektuře, neboť jednotlivá řešení se mohou značně lišit.

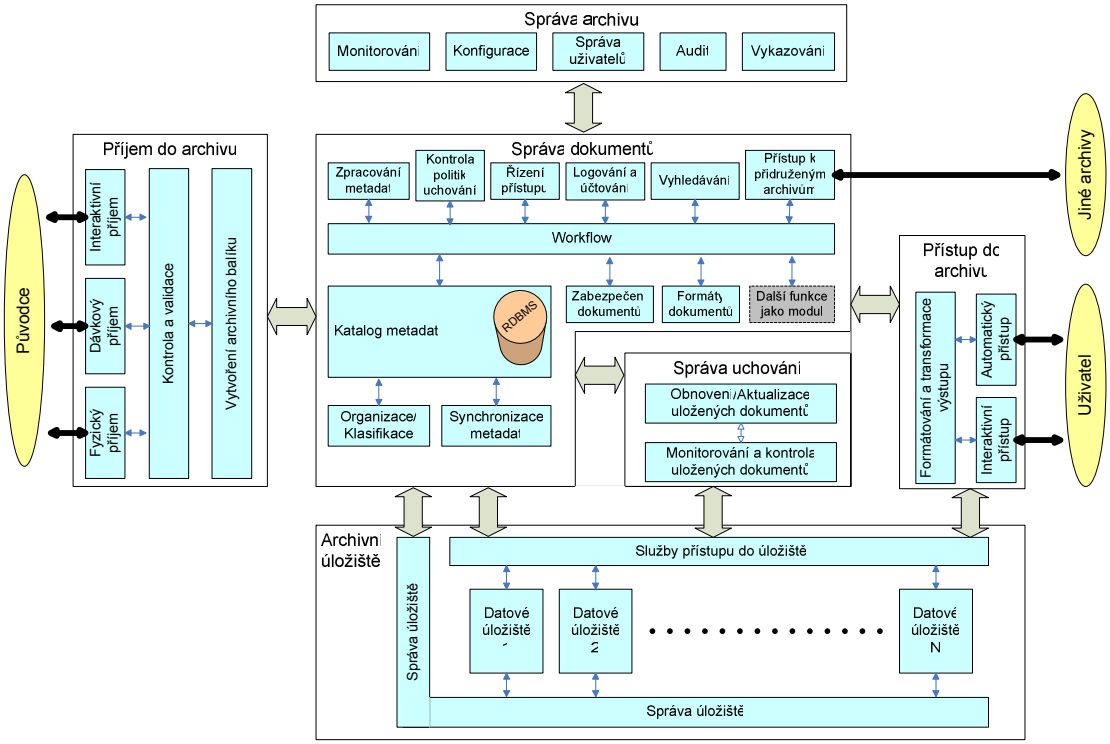
Přesto na obecné úrovni lze ukázat schéma architektury, která může sloužit jako rámcové východisko pro implementaci archivu. Budeme zde především vycházet ze standardu OAIS a ze zkušeností některých současných projektů[56], vycházejících z tohoto standardu.

Hlavními požadavky na architekturu, které vyplývají z předchozích analýz jsou:

- škálovatelnost – architektura musí podporovat plynulý růst kapacity a výkonu systému (u digitálních archivů lze často předpokládat i exponenciální růst objemu uložených dat)
- flexibilita a dynamičnost – tak, aby se archiv mohl rozvíjet s příchodem nových technologií a nahrazovat zastaralé
- spolehlivost a dostupnost– systém musí poskytovat vysokou spolehlivost pro uložení dokumentů i pro zajištění přístupu k nim. Jak bylo ukázáno v předchozím textu, redundance úložišť i ostatních komponent je jedním z hlavních prostředků jak toho dosáhnout
- bezpečnost – archiv musí zajistit bezpečné uložení dokumentů po celou dobu jejich životnosti
- podpora heterogenních technologií – z dlouhodobého hlediska je zřejmé, že archiv bude obsahovat různorodé technologie, architektura proto musí obsahovat vhodnou integrační infrastrukturu, která zajistí fungování systému jako celku.

[56] FEDORA – otevřený archivační systém, http://www.fedora.info/;
NEDLIB – project EU http://nedlib.kb.nl/;
NARA project http://www.sdsc.edu/NARA;
DSpace otevřený archivační systém http://dspace.org/;
British Library - The large-scale archival storage of digital objects, Jim Linden, Sean Martin, Richard Masters, and Roderic Parker, The British Library,DPC Technology Watch Series, Report 04-03, February 2005

Schéma na doprovodném obrázku ukazuje rámcovou architekturu archivu.



Jádrem řešení je **systém pro správu dokumentů**[57], který zajistí řízení životního cyklu elektronického dokumentu v archivu a poskytuje základní funkcionalitu potřebnou pro práci s dokumentem a archivním balíkem.
Tento systém jednak zajišťuje infrastrukturu pro práci s dokumenty a oběh dokumentů (resp. archivního balíku), kde modul **Workflow** definuje tok dokumentů a potřebné procesy pro jednotlivé úkoly.

Klíčovým modulem je **Katalog metadat**, což je vlastně databáze, která obsahuje doplňující informace o uložených dokumentech potřebné jak pro vyhledávání, tak pro správu archivovaných dokumentů. Spolehlivost a dostupnost tohoto modulu je kritická pro celý archiv, proto musí být zajištěna pomocí vhodných technologií jako jsou cluster, on-line replikace, automatické zálohování aj.

Katalog metadat může využívat další moduly, které přispívají ke správě metadat:

- **Organizace/Klasifikace** – umožní klasifikovat dokumenty do vhodného systému tak, aby bylo možné snadno vyhledávat relevantní informace v archivu. Klasifikace se provádí na základě metadat dokumentu případně i obsahu dokumentu a může probíhat automaticky s využitím sofistikovaných třídících algoritmů a systému strojového zpracování textu nebo manuálně za asistence personálu archivu a nebo vhodnou kombinací obou přístupů.
- **Synchronizace metadat** – provádí synchronizaci metadat uložených v databázi s metadaty uloženými v archivním úložišti a propaguje případné změny mezi oběma subsystémy.

[57] obecněji můžeme říci systém pro správu dat, jak např. uvádí OAIS

Systém správy dokumentů dále musí obsahovat modul **Řízení přístupu**, který zajišťuje autorizovaný přístup k dokumentům ve všech krocích zpracování, a **Logování a účtování**, který zaznamenává všechny akce nad dokumenty do logů.

Dále obsahuje systém správy dokumentů jednotlivé funkční moduly, které poskytují specifické služby v různých krocích zpracování elektronického dokumentu:

- **Zpracování metadat –** modul sloužící pro vytváření, čtení a kontrolu metadat struktur. Součástí jsou i definice schémat a slovníků metadat. Je důležité, aby tento modul byl dostatečně otevřený a flexibilní, aby umožnil definici prakticky libovolné sady metadat a její budoucí rozšiřování.
- **Kontrola politik uchování –** modul umožňuje definici politik pro uchování dokumentu a poskytuje kontrolu dokumentu a archivního balíku proti těmto politikám.
- **Vyhledávání –** modul umožňuje vyhledávání dokumentů podle různých atributů, zpracování komplexních výrazů pro vyhledávání atd.
- **Přístup k přidruženým archivům –** modul umožňuje přístup k metadatům a dokumentům v jiných archivech. Využívá standardní protokoly pro komunikaci mezi archivy jako jsou OAI-MHP[58] nebo OpenURL[59].
- **Zabezpečení dokumentů –** modul zajišťuje různé bezpečnostní služby pro dokumenty a archivní balíky, týká se jak ověřování, tak vytváření bezpečnostních prvků (otisky, podpisy, atd.).
- **Formáty dokumentů –** modul slouží pro identifikování formátů, ověřování validity formátů a případné konverze formátů.

Kromě definovaných funkčních modulů, které jsou zde popsány, by měl systém správy dokumentů umožnit přidání dalších modulů, které mohou vzniknout v budoucnu pro pokrytí dalších nových požadavků a jejich začlenění do procesu zpracování archivovaných dokumentů.

Součástí správy dokumentů je také specifický subsystém **Správy uchování**, který plní specifické služby z hlediska dlouhodobého uchování dokumentů:

- modul **Monitorování a kontroly uložených dokumentů** prochází uložené dokumenty a kontroluje jejich integritu a shodu jejich vlastností s politikami archivu a označuje dokumenty, které vyžadují nějaké akce, aby byla zachována jejich použitelnost.
- modul **Obnovení/aktualizace uložených dokumentů** umožňuje aktualizovat uložené dokumenty v souladu s politikami uložení tak, aby byly nadále přístupné a použitelné. Jedná se zde především o možné migrace formátu, úpravy metadat nebo i případné aktualizace zabezpečení dokumentů.

Dokumenty spolu s metadaty jsou uloženy v **Archivním úložišti**, které má za úkol zaručit přesné uchování binární reprezentace archivního balíku po celou dobu uložení a zajistit k ní přístup kdykoliv je potřeba.

[58] The Open Archives Initiative Protocol for Metadata Harvesting, Protocol Version 2.0 of 2002-06-14, http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm
[59] ANSI/NISO Z39.88 -2004, The OpenURL Framework for Context-Sensitive Services

Archivní úložiště je postaveno na technologiích datových skladů a kromě vysoké bezpečnosti a spolehlivosti, musí být zaručena i jeho vysoká škálovatelnost. Jak již bylo uvedeno v kapitole 2.3.4.1, jedním z hlavních prostředků pro zajištění výše uvedeného je replikace uložených dat a využití několika typů datových nosičů. Typické archivní úložiště bude tedy obsahovat několik **Datových úložišť**, která mohou být i v různých lokacích a využívat rozdílné technologie. S nárůstem archivu budou přibývat další úložiště a případně nové systémy budou nahrazovat staré. Proto, aby je bylo možné jednotně spravovat, definovat mechanismy replikací, přidávat nová úložiště atd., je potřebné implementovat modul **Správy úložiště**, jako je například systém HSM (Hierarchical Storage Management), který byl popsán v kapitole 2.3.4.2. Z předchozího popisu je zřejmé, že úložiště dat budou tvořit heterogenní systémy, které se budou měnit a rozvíjet. Pro zajištění nezávislosti na změnách v archivním úložišti pro ostatní částí archivu je potřebné implementovat modul **Služby přístupu do úložiště**, který poskytuje vhodnou abstrakci přístupu k uloženým datům a definuje jednotné rozhraní a poskytuje další potřebné služby. Příkladem takového systému může být např. SRB[60], který poskytuje tyto služby pro vysoce distribuované úložiště s velkými objemy dat.

Dalšími částmi archivního řešení jsou systémy pro Příjem do archivu a Přístup do archivu, které tvoří rozhraní archivu, kde první je určen pro původce dokumentů a druhý pro uživatele, kteří potřebují získat informace z archivu.

Modul **Příjem do Archivu** obsahuje především moduly rozhraní, které umožňují různé způsoby předání dokumentů do archivu:

- **Interaktivní příjem** – elektronický příjem jednotlivých dokumentů s tím, že se předpokládá určitá interakce s původcem – vyplnění online formulářů aj.,
- **Dávkový příjem –** určený pro automatický příjem dokumentů (i většího množství najednou),
- **Fyzický příjem –** kdy původce fyzicky doručí dokumenty do archivu.

Jednotlivé příjmové moduly také zajistí převedení do jednotného formátu příjmového balíku, který je dále zpracován modulem **Kontroly a Validace**, kde jsou ověřeny všechny jeho potřebné vlastnosti, ať již se týká formátu, zabezpečení nebo dostupných metadat. Pokud jsou kontroly kladné a příjmový balík splňuje požadavky archivu, jsou v modulu **Vytvoření archivního balíku** přijaté dokumenty a metadata upraveny do standardního archivního balíku a předány k uložení.

Modul **Přistup do archivu** umožňuje zase oprávněným uživatelům získat dokument a/nebo jeho metadata. Přistup může být buď interaktivní (který zajišťuje modul **Interaktivní přístup**), pomocí vhodného uživatelského rozhraní, nebo archiv umožní automatický přístup pro jiné systémy (modul **Automatický přistup**), pomocí vhodných standardních protokolů (např. výše jmenované OAI-MHP, OpenURL aj.)

Dále také archivní systém obsahuje moduly pro **Správu Archivu**, které zajišťují uživatelské rozhraní pro:

- **Konfiguraci** celého systému
- **Správu uživatelů** jejich vlastností, rolí a přístupů
- **Monitorování** funkcí systému

[60] SDSC Storage Resource Broker, http://www.sdsc.edu/srb/index.php/Main_Page

- **Vykazování** různých souhrnných hlášení o uložených dokumentech i o procesech v systému
- **Audit –** poskytování informací pro audit archivu

Jak vyplývá z výše uvedeného schématu, systém digitálního archivu obsahuje řadu komponent. Je málo pravděpodobné a taky nežádoucí z hlediska flexibility a budoucího rozvoje systému, že by všechny komponenty tvořily jednu monolitní aplikaci. Proto základem řešení musí být vhodný rámec, který umožní integraci všech komponent systému do funkčního řešení.

Vhodným přístupem, který vychází ze současných trendů a který se také uplatňuje v řadě projektů archivačních systémů, je SOA – Service Oriented Architecture[61] (mluví se také často o kompozitní architektuře). Jedná se o princip návrhu aplikace, kdy místo jednolité aplikace je řešení postaveno jako sada relativně nezávislých komponent, které poskytují dobře definované služby a které jsou volně propojené do jednoho funkčního celku. Tento přístup umožňuje dynamicky přidávat nové služby a rozšiřovat a měnit stávající funkcionalitu a zajistit tak evoluci systému spolu s měnícími se požadavky (a jak bylo ukázáno výše toto je jeden z hlavních imperativů pro systém archivu). Pro architekturu kompozitního systému vyplývají především tyto požadavky:

- Funkcionalita systému je zabalena do komponent, které poskytují dobře definované služby.
- Služba má přesně definovaný kontrakt, který popisuje, jak přesně služba má být použita (rozhraní, protokol).
- Abstrakce služby – vše co je potřebné pro využití služby je popsáno v jejím kontraktu. Není nutné znát detaily implementace.
- Služby jsou autonomní – obsahují vše co potřebují pro svoji funkci.
- Jednoduché služby lze skládat do složitějších, komplexních.
- Služby jsou volně spojené tak, aby poskytovaly potřebnou funkcionalitu aplikace.
- Služby mohou být znovu využity v různých kontextech.
- Služby mohou být nalezeny podle potřeby.

Aby řešení mohlo fungovat na SOA principech potřebuje společnou infrastrukturu, která poskytuje obecné funkce systému jako komunikaci mezi službami, registraci služeb, kompozici složených služeb, workflow atd. Tato infrastruktura, obecně nazývaná middleware, slouží jako výchozí platforma, na které je potom budována vlastní logika aplikace jako sada služeb a jejich vzájemných interakcí. Na trhu existuje řada firem, které nabízejí middleware vhodný pro vybudování rozsáhlých kompozitních aplikací[62]. Dobrou praxí bývá využití kvalitní dostupné platformy middleware, neboť se může jednat o poměrně komplexní systém a jeho chování ovlivňuje klíčové vlastnosti výsledného řešení jako je stabilita a výkonnost. Vychází se z předpokladu, že specializované firmy jsou schopny se zhostit tohoto úkolu

[61] Obecný popis přístupu k SOA je popsán v Reference Model for Service Oriented Architecture 1.0, Committee Specification 1, 2 August 2006, OASIS, http://www.oasis-open.org/committees/download.php/19679/soa-rm-cs.pdf

[62] Např. Tibco, Oracle, BEA, WebMethods, Sun aj.

nejlépe. Kvalitní middleware poskytne také vhodné východisko pro architekturu jednotlivých komponent systému, které musí být vyvinuty.

Moderní systémy správy dokumentů (o kterých jsme mluvili výše) typicky obsahují nějaký SOA middleware, ať již obecný nebo specifický pro danou aplikaci. Pokud tedy archiv je budovám na nějakém takovémto systému, lze samozřejmě využít tuto architekturu a pomocí ní zajistit potřebné rozšíření/změny, které jsou potřebné pro archivaci.

Dostupné systémy pro archivaci se snaží poskytnout některé funkce popsané výše, ale jak již bylo uvedeno na začátku této kapitoly, neexistuje jediné řešení, které by pokrylo všechny požadavky. Většina nabízených řešení vychází z podnikových systémů správy dokumentů[63] (ostatní případně ze systémů správy uložení dat nebo systémů automatizace knihoven[64]) , které jsou doplněny o některé archivní funkce. Ovšem funkce specifické pro dlouhodobé uchování nejsou většinou dostupné, proto si je řeší archivy sami jako doplňující moduly.
Existují také nové otevřené projekty (FEDORA[65], DSpace[66]), které se snaží vytvořit platformy speciálně navržené pro zajištění dlouhodobého uchování. Nicméně i zde, kde se jedná o aplikaci posledního výzkumu a vývoje v této oblasti, některá funkcionalita není zatím dostupná (např. FEDORA má moduly pro správu uchování zatím ve stadiu přípravy návrhu).

2.4 Organizační zabezpečení archivu

Pro vybudování digitálního archivu, který by převzal náležitou a dlohodobou péči o digitální fondy, musí organizace dát dohromady a sladit několik klíčových komponent. Jsou to:

- organizace a organizační infrastruktura
- procedury, procesy a funkce archivu
- technologická infrastruktura

Pro to jak mají tyto komponenty vypadat je samozřejmě klíčová podoba digitálních fondů, o které bude digitální archiv pečovat.

2.4.1 Digitální fondy

Při projektování digitálního archívu je třeba vzít do úvahy charakter digitálních dokumentů, které archiv bude spravovat. Při tom je nutné vzít v úvahu následující faktory:

Rozsah a měřítko: velikost sbírek a jejich počet, počet souborů, velikost souborů;
Komplexnost: budou sbírky homogenní nebo heterogenní, digitální dokumenty budou jednoduché anebo komplexní;
Hodnota: hodnota ukládaných dokumentů a sbírek ve vztahu k postavení a poslání archivu;

[63] OpenText, EMC Dokumentum a další
[64] Exlibris
[65] FEDORA – otevřený archivační systém, http://www.fedora.info/
[66] DSpace otevřený archivační systém http://dspace.org/

Kontrola: úroveň a charakter kontroly nad přijímanými a spravovanými dokumenty včetně dlouhodobého přístupu k nim.

Zdroje určené pro založení archivu a jeho provoz musí být v souladu s výše uvedenými faktory.

2.4.2 Procedury, procesy a funkce archivu

Tato část se bude zabývat procedurami, procesy a funkcemi potřebnými pro přijímání, spravování a zajištění přístupu k elektronickým dokumentům v dlouhodobém časovém horizontu. Požadavky na technické zajištění a potřebnou infrastrukturu nejsou v této části řešeny a je jim věnována samostatná část (viz 2.3.6). Požadavky pro výše zmíněné činnosti jsou rozděleny do pěti částí.

První část shrnuje potřeby související s vkládáním a přijímáním digitálního obsahu. Vkládání je klíčovým bodem součinnosti mezi archivem a původcem. Z hlediska archivu je to klíčový moment pro získání potřebné kontroly nad vkládaným a spravovaným dokumentem.
Druhá část stanovuje minimální podmínky pro dlouhodobé uchování dokumentů (ve formě archivních balíků – AIP). Systémová infrastruktura musí poskytovat služby umožňující pracovat s AIP.
Třetí část se zabývá definicí strategií, které musí mít archiv implementovány, aby byl dlouhodobě zajištěn přístup k jeho obsahu.
Část čtvrtá se zabývá požadavky na minimální množinu metadat potřebnou k tomu, aby elektronický dokument mohl být archívem spravován a též v archivu nalezen.
Nakonec část pátá se zabývá problematikou přístupu k uloženým dokumentům a otázkami souvisejícími s přesností a autentičností poskytovaných informací.

2.4.2.1 Vkládání a přijímání obsahu

„Vložení“ je obecný pojem popisující proces, který se odehrává předtím, než elektronický dokument se stane objektem uloženým v archivu. V závislosti na poslání archivu, povaze uchovávaných dokumentů, vztahů archivu s původci se proces vložení bude lišit archiv od archivu. Obecně lze říci, že proces vložení je ukončen v okamžiku kdy AIP a s ním asociovaná metadata jsou bezpečně uložena v archivu, včetně vytvoření bezpečností kopie. Obtížněji se určuje, kdy proces vložení začíná a hodně se odlišuje v závislosti na vztazích archiv-původce. Tyto vztahy se mohou značně odlišovat stupněm formálnosti a vymezením vzájemných povinností.
Vzhledem k těmto rozdílnostem, lze požadavky na proces vložení formulovat pouze velmi obecně a bude záležet na posouzení jednotlivého archivu co bude vyžadovat. Toto posouzení se bude zejména odvíjet od poslání archivu a od potřeb cílových uživatelů a proto i původci musí být zcela jasný vztah mezi archivem, jeho posláním, jeho sbírkami a cílovými uživateli. Touto logikou se potom musí i řídit doplňující informace asociované s primárními elektronickými dokumenty archivu. Informace, které budou vloženy spolu s primárními elektronickými dokumenty, budou formálně vymezeny. Tyto informace musí umožnit cílovému uživateli smysluplně a s pochopením použít uložený dokument, aniž by komunikoval s jeho původcem, expertním pracovníkem archivu anebo jiným expertem.
Jak již bylo řečeno, je v obecnosti vyloučené vyjmenovat dokumentaci požadovanou k elektronickému dokumentu uloženému v certifikovaném archivu. Taková

dokumentace bude obsahovat metadata, příklady formulářů, kontextové informace, odkazy na související dokumenty, související studie, atd.. Dokumentace se vytváří jednak proto, aby umožnila správnou interpretaci dokumentu a jednak aby uživateli umožnila posoudit přesnost a správnost dokumentem poskytovaných informací.
V referenčním OAIS modelu je přístup následující. Základním úkolem archivu je uchovávání informací, tzn. elektronických dokumentů společně s jejich Reprezentačníma informacemi. To jest primární informace, která má být uchována a nazývá se v OAIS terminologii Obsahová informace. Základní rozhodnutí, které musí archiv společně s původcem učinit, je definovat co tvoří informaci, která má být uložena, neboli Obsahovou informaci. Doporučení OAIS je začít rozhodnutím, co je primární elektronický dokument a pak určit co má být obsaženo v Reprezentační informaci, která bude tento elektronický dokument doprovázet.

Následující příklady ukazují typ a rozsah dokumentace pro různé datové typy a archivy.

a) Pro každou třídu elektronických dokumentů archiv jasně identifikuje a deklaruje vlastnosti, které bude zachovávat.
b) S každým původcem uzavře archiv písemnou smlouvu, která bude vymezovat všechny aspekty vkládání, obhospodařování, přístupu a vyjímání dokumentů.
c) Pro každou třídu elektronických dokumentů má archiv písemně stanovenou jejich formu pro vkládání, doplňující dokumentaci, a omezení přístupu.
d) Archiv má implementovaný proces ověřující, že získaná informace pochází z předpokládaného zdroje.
e) Nad dokumenty uloženými v archivu má archiv dostatečnou fyzickou kontrolu. Archiv může získat fyzickou kontrolu nad elektronickými dokumenty některou z následujících činností:
    - Analýzou digitálního obsahu
    - Verifikací, analýzou a vytvořením metadat
    - Autentizací a kontrolou integrity
    - Vytvořením Archivního Informačního Balíčku (AIP) – encapsulizace
f) Během procesu vložení archiv ověří, že každý SIP je úplný a korektní.
g) Během procesu vkládání archiv poskytuje původci informace o postupu jednotlivých kroků.
h) Archiv může zdokumentovatelným procesem ukázat, že všechny informace poskytnuty v rámci SIP jsou buď zahrnuty do výsledného AIP anebo věrohodným a ověřitelným způsobem zničeny.
i) Archiv závazně informuje původce o době ukončení procesu vložení a převzetí odpovědnosti za SIP archivem.

2.4.2.2 Správa uložených informací

a) Pro každý AIP nebo každou třídu elektronických dokumentů uchovávanou archivem má archiv napsanou vymezující definici.
b) Pro každý AIP (nebo třídu AIP) má archív definici adekvátní potřebám v souvislosti s dlouhodobým uložením.
c) Archiv má definováno jakým způsobem se ze SIP vytvoří AIP.
d) Archiv má a používá pojmenovávací konvenci, která poskytuje viditelné a jednoznačné identifikátory pro všechna AIP.

e) Pokud je jednoznačný identifikátor asociován se SIP před jeho vložením, je tento zachován a asociován s výsledným AIP.
f) Pro každé vygenerované AIP ověří archiv úplnost a správnost.
g) Archiv má implementován mechanismus, který umožňuje nezávislý audit integrity sbírek a obsahu sbírek.

2.4.2.3 Plánování uchování, migrace a další strategie

Pro důvěryhodný archiv nestačí pouze uchovávat informaci. Důvěryhodný archiv musí uchování provádět v souladu s jasně definovanou, zdokumentovanou strategií s použitím předem odsouhlasených mechanismů a procedur. Pokud archiv nevyhoví těmto požadavkům, nelze ho považovat za důvěryhodný a nemůže projít žádným certifikačním auditem.
Výše zmiňovaná dokumentace nemusí být nijak složitá a komplexní. Nemusí též detailně popisovat, jak se archiv vypořádá s neznámými věcmi, které vyvstanou v budoucnu. Např. je zcela v pořádku pokud archiv nedokumentuje jak bude uchovávat soubor ve formátu, který ještě neexistuje. Nicméně ve strategii archivu by mělo být popsáno co archiv udělá, pokud mu někdy bude předložen soubor ve formátu, se kterým se doposud nesetkal. Popsaná politika může být odmítnutí, rozhodnutí na zhodnocení využitelnosti, atd.

Důvěryhodný digitální archiv musí mít písemně, jasně a transparentně stanovenou strategii, zásady, procedury a procesy. Tato dokumentace musí být explicitní, srozumitelná, nedvojznačná, úplná, aktuální a obecně dostupná.

Archiv musí být schopný demonstrovat:
- Jasné stanovisko k přijímaným formátům – tj. jasně formulované zásady, které definují, vymezují a omezují přijímané formáty.
- Srozumitelný postup při přijímání elektronických dokumentů; příklady – protokoly o převodu včetně úloh a zodpovědností původce a archivu, dokumentaci konverzí při generováni AIP, mechanismy pro kontrolu kvality.
- Uskutečněné a plánované konzervační akce, strategii konzervování, užité metodiky, logy z uskutečněných akcí.
- Archivní strategii, procesy a praxi pro zajištění efektivního přijímání, průběžného a budoucího uložení, včetně anticipace nevyhnutelných technologických změn.
- Nezávislé prostředky, založené na registrování přijímaných dokumentů, na ověření obsahu archivu – např. auditovatelný registr přijímaných dokumentů.

Plánování uchování musí též obsahovat strategii jak archiv bude provádět své klíčové aktivity v měnícím se okolním prostředí (technologickém, sociálním, atd.). Tato strategie musí řešit:
- Proces monitorování změn, které mohou ovlivnit dlouhodobé uchovávání.
- Zajištění expertní analýzy a interpretace dopadu těchto změn.
- Plánovanou odpověď na tyto změny.
- Implementaci odpovědi.

Strategie uchování se též musí zabývat otázkou, za jakých podmínek umožnit vymazání AIP z archivu. Například archiv drží určité informace/dokumenty ve

formátu, pro který vhodné software přestávají být podporovány a postupně se stanou nepoužitelné. Možné strategie jsou:

- Provést potřebnou transformaci dat při vkládání
- Zůstat u původní formátu a čekat, že někdo jiný problém vyřeší.
- Vytvořit emulační prostředí, ve kterém software bude dále funkční.

Je možné, že určení vhodné strategie bude nutné pro každou třídu (tj. formát) digitálních dat v archivu uložených.

Další strategie je nutná pro další práci s AIP a kontroly nad AIP mimo běžných kontrol v rámci systémové robustnosti. Během uchování v archivu bude AIP pravděpodobně doplňován o další PDI (Preservation Description Information), tudíž budou vznikat nové verze AIP. Kromě toho na SIP a AIP mohou být aplikovány další transformace, které opět budou generovat nové verze AIP. Jako příklad uveďme transformaci dokumentu do nového, snadněji podporovaného formátu. Z hlediska důvěryhodnosti a auditovatelnosti je důležité, aby aktualizované záznamy (logy, historie) těchto změn byla uchovávána, včetně přijatého SIPu a vytvořených verzí AIP. Optimální verzi detailu je těžké stanovit. Za postačující lze pravděpodobně označit zachování takového detailu, který umožní zpětné vytvoření jedné verze z druhé v průběhu celé historie uchovávaného dokumentu.

Pro důvěryhodný digitální archiv by mělo platit zejména:

a) Archiv má aktuální dokument popisující strategii uchování.
b) Archiv má implementovány podstatné části strategie uchování.
c) Archiv užívá vhodné mezinárodní registry pro reprezentaci informací (Representation information registry). Global Digital Format Registry (GDFR), britský registr formátů v rámci Národního archivu PRONOM a britský Digital Curation Centre's Representation Information Registry jsou tři příklady nově vznikajících mezinárodních standardů, které archivy mohou použít. Archivy by měly, kdykoliv je to možné, užívat tyto standardy pro namapování Reprezentačních informací z Obsahové informace a PDI.
d) Archiv uchovává Reprezentační informace (včetně formátů) pro všechny vložené dokumenty.
e) Archiv uchovává a konzervuje Obsahové informace v rámci AIP.
f) Pro každou Obsahovou informaci archiv vytvoří, udržuje a uchovává aktuální PDI. PDI je potřebné nejenom v rámci práce archivu pro kontrolu neporušenosti Obsahové informace a vyhledávání, ale hlavně tím, že obsahuje informace o původu a kontextové informace pro zajištění adekvátního porozumění Obsahové informaci. Detail a rozsah kontextových informací a informací o původu je dán potřebami cílových uživatelů.
g) Archiv aktivně monitoruje integritu uložených AIP.
h) Archiv udržuje aktuální záznamy o svých aktivitách spojených s vkládáním, archivačními procesy a s těmi administrativními procesy, které souvisí s archivací.
i) Archiv má implementovaný mechanismus, který sleduje stav Reprezentační informace (tj. formátů) a poskytuje varování, pokud tato informace začíná zastarávat a z tohoto důvodu přestává být použitelná.
j) Pokud v rámci monitorování činnosti archivu vyvstane potřeba provést změny v Plánu uchování má archiv implementovány procesy na provedení těchto změn.

k) Archiv může dokumentovat úspěšnost své Strategie uchování.

2.4.2.4 Správa metadat

Kritickou částí každého archivu je jeho funkcionalita v oblasti správy dat. Bez ohledu na technické detaily, systém musí být schopen ukládat deskriptivní informace (metadata) pro zajištění přístupu k obsahu sbírek a vyhledávání.

a) Archiv přijímá metadata anebo vytváří minimální metadata a zajišťuje jejich spojení s AIP.
b) Archiv musí zajistit a udržovat aktuální referenční integritu mezi všemi AIP a asociovanými deskriptivními informacemi.

2.4.2.5 Řízení přístupu

Součástí celkové politiky archivu je politika týkající se přístupových práv.

a) Archiv musí mít plně implementovanou aktuální přístupovou politiku.
b) Archiv vede záznam o všech selhání přístupové politiky a všechny incidenty typu „Access denial“ jsou prověřovány.
c) Poskytování informací uživateli je zorganizováno tak, aby uživatel byl vždy přesně a vyčerpávajícím způsobem informován o tom, co dostal. Jinými slovy, proces, který generuje DIP (Dissemination Information Package) je úplný a řádně ukončený vzhledem k požadované informaci. To znamená, že následující scénáře jsou nepřípustné:
    - Požadavek lze uspokojit pouze částečně a je vygenerován neúplný DIP, ale odpověď doručená uživateli neindikuje neúplnost.
    - Požadavek je odložen na neurčito, protože něco z toho, co požaduje, není momentálně dostupné. Uživatel není informován nebo není informován, kdy bude informace dostupná.

d) Všechny žádosti o přístup musí být vyřešeny – buď povolením nebo zamítnutím.
e) Archiv musí být schopen vydávat buď autentické kopie originálu anebo dokumenty s doložitelným vztahem vzhledem k originálu.

2.4.3 Organizace a organizační zabezpečení

Odpovídající technologie a procesy tvoří pouze jeden ze základů TDR. Druhým, neméně důležitým, je organizační zabezpečení. Řádné organizační zabezpečení musí splňovat řadu kritérií v mnoha oblastech. Tyto oblasti jsou následující:

1. Řízení a organizační schůdnost
2. Organizační struktura a personální obsazení
3. Kontrolovatelnost procesů a provozní řád
4. Finanční udržitelnost
5. Kontrakty, licence a závazky

2.4.3.1 Řízení a organizační schůdnost

a) Archív musí jasně deklarovat své poslání (mission statement), v němž je jasně deklarován jeho závazek a odpovědnost za dlouhodobé uchování, zprávu a zpřístupňování elektronických dokumentů ve prospěch a v zájmu původců.
b) Pro případ nepředvídaných událostí, pro případy, že archív z nějakého důvodu musí omezit anebo zrušit své fungování má archiv naformulovány a implementovány plány pro případného následníka (succession plan), plány pro nepředvídané situace (contingency plans) a eventuelně též plány pro případné dočasné uložení.

2.4.3.2 Organizační struktura a personální obsazení

a) Zaměstnanci archívu mají znalosti a zkušenosti odpovídající jejich úkolům.
b) Archiv disponuje dostatečným počtem zaměstnanců k řádnému zajištění všech funkcí a služeb dohodnutých a slíbených původcům a uživatelům.
c) Archív jednoznačně deklaruje odhodlání podporovat a organizovat další vzdělávání zaměstnanců za účelem udržení a zvýšení jejich znalostí a zkušeností.

2.4.3.3 Kontrolovatelnost procesů a provozní řád

a) Archív má implementovány mechanismy pro průběžné monitorování, zdokonalování a vývoj nových procesů a politik, které odrážejí potřeby vyplývající z růstu archívu a vývoje a růstu potřeb obsluhované komunity.
b) Pro zajištění kontinuity fungování archivu, včasné řešení problémů či požadavků klientů implementuje archiv systém, který monitoruje a reaguje na požadavky a problémy.
c) Archiv se zaváže k pravidelným, oficiálním (formálním) přezkoumáním a zhodnocením, aby byl podpořen další, nepřetržitý vývoj.
d) Archív vede podrobný záznam o všech aktivitách a změnách vztahujících se jeho fungování, operativně, procesům, software, hardware.
e) Archív se formálně zaváže k průhlednosti a zodpovědnosti při všech aktivitách ovlivňujících operativu a řízení archívu.
f) Archív se formálně zaváže, že bude definovat a implementovat ukazatele pravdivě zachycující kvalitu a integritu jím poskytovaných služeb. Dále se zaváže, že bude tyto ukazatele sledovat, shromažďovat a na požádání je poskytne.
g) Archív se formálně zaváže k pravidelným recertifikacím. Kromě toho se archív zaváže neprodleně informovat vydavatele certifikátu o všech operačních změnách, které by mohly jeho certifikát změnit anebo anulovat.

2.4.3.4 Finanční stabilita

a) Archív implementuje krátkodobé i dlouhodobé finanční plánování na podporu své finanční stability.
b) Archív implementuje formální vyhodnocovací a revizní proces jeho finanční situace, který dle potřeby upraví jeho finanční plány. Toto vyhodnocení a revize se koná nejméně jednou za rok.

c) Archív vede transparentní finanční účetnictví, odpovídající současné účetní praxi a ve shodě s relevantním účetními předpisy a standardy. Finanční hospodaření archívu je auditováno.
d) Archív se formálně zaváže průběžně a pravidelně analyzovat své pohledávky a závazky, licence, rizika, benefity, výdaje a investice a tyto analýzy pravidelně zveřejňovat.

2.4.3.5 Kontrakty, licence a závazky

a) Jestliže archív spravuje, uchovává a/nebo poskytuje přístup k elektronickým dokumentům v zastoupení jiné organizace, tak tyto služby poskytuje na základě platných smluv.
b) Pokud archív uchovává a spravuje elektronický dokument pro třetí stranu a na základě smlouvy, musí tato přesně specifikovat převod práv a povinností spojených s uchováváním daného elektronického dokumentu.
c) Archív sleduje a řídí aplikaci copyrightu a dalších omezení použitelnosti uložených dokumentů tak, jak je požadováno licencemi a smlouvami.
d) Archív má vytvořenou politiku (vnitřní směrnici) omezující potenciální zodpovědnost a rizika související s ochrannou duševního vlastnictví, pro případ přijmutí elektronického dokumentu s nevyjasněnými majetkovými anebo autorskými právy.

2.4.4 Plán obnovy a krizové řízení

Plány obnovy a krizového řízení jsou v podstatě velmi podobné pro všechny informační systémy a existuje mnoho literatury, která se jejich vývojem, implementací a údržbou zabývá.. Vzhledem k tomu je v následujících odstavcích probereme spíše stručněji a pro podrobnější informace uvedeme relevantní odkazy.

Proces plánování obnovy a krizového řízení má sedm kroků:

a) Vytvoření koncepce a definice zásad pro Plán obnovy a krizového řízení
b) Provedení analýzy obchodních rizik (Business impact analysis)
c) Návrh preventivních opatření a kontrol
d) Vytvoření strategie obnovy
e) Vytvoření plánů obnovy IT
f) Vytvoření plánů testování, výcviku a cvičení
g) Implementace a údržba plánu

2.4.4.1 Vytvoření koncepce a definice zásad pro Plán obnovy a krizového řízení

Pro efektivní implementaci a realizaci plánu obnovy je nezbytné, aby všichni zaměstnanci archívu znali a správně chápali cíle a požadavky plánu obnovy. K tomu je nutné jasně a zřetelně stanovit zásady obnovy archívu. Koncepce plánu obnovy musí jasně stanovovat celkové cíle při obnově a formulovat organizační rámec a zodpovědnosti v rámci plánu obnovy. Nezbytnou podmínkou úspěchu je podpora vyššího managementu, zejména CIO. Vyšší management by měl aktivně participovat na tvorbě koncepce programu, struktuře, definicích cílů a rozdělení úloh a zodpovědností. Klíčové prvky koncepce jsou:

- Úlohy a zodpovědnosti

- Vymezení rozsahu a odpovědností vzhledem k jednotlivým platformám a organizačním funkcím (útvarům)
- Požadavky na zdroje
- Testovací a cvičné plány
- Rozvrh údržby plánu obnovy
- Rozvrh zálohování záložních médií

Vývoj plánů obnovy je nutné koordinovat s dalšími činnostmi archivu, včetně IT bezpečnosti, fyzické bezpečnosti, lidskými zdroji, atd..

2.4.4.2 Provedení analýzy obchodních rizik (Business impact analysis)

Analýza obchodních rizik umožní koordinátorovi plánu obnovy úplně popsat a charakterizovat systémové požadavky, procesy a vzájemné závislosti a na základě těchto informací určit požadavky a priority pro mimořádné situace. Dalším cílem analýzy je nalézt závislosti mezi kritickými službami a specifickými systémovými součástmi potřebnými pro jejich zajištění a na základě těchto znalostí vyhodnotit důsledky poruch těchto částí systému. Klíčové kroky analýzy jsou:

- Identifikace kritických součástí IT
- Analýza důsledků omezení funkčnosti/výpadku části systému a tolerovatelné doby poruch
- Stanovení priorit obnovy

2.4.4.3 Návrh preventivních opatření a kontrol

V některých případech lze důsledky výpadků a/nebo poruch eliminovat nebo omezit pomocí preventivních opatření, která zamezují nebo snižují dopady na systém. Použití preventivních metod lze doporučit všude tam, kde je to možné a cenově výhodné. Existují celé škály preventivních kontrol podle konkrétního typu systému a jeho konfigurace. Uvedeme proto jen pár ilustrativních, jež lze aplikovat pro všechny:

- Krátkodobé i dlouhodobé záložní zdroje elektrické energie
- Detektory kouře a ohně
- Hasicí systém
- Detektory vody ve stropech a podlahách počítačových sálů
- Plastové plachty na ochranu elektronických zařízení proti vodě
- Ohnivzdorné a vodotěsné kontejnery na záložní média a kritické nedigitální záznamy
- Centrální bezpečnostní vypínač
- Sekundární úložiště umístěné v jiné lokalitě pro uchovávání záloh, systémové dokumentace a klíčových nedigitálních záznamů
- Kontroly bezpečnostních opatření

Preventivní kontroly je nutné průběžně dokumentovat v plánu pro mimořádné situace.

2.4.4.4 Strategie obnovy

Účelem strategie obnovy je vytvoření podmínek a prostředků pro rychlé obnovení funkčnosti archívu.Strategie musí vycházet z analýz důsledků výpadků částí systému a tolerovatelné doby poruch. Na strategii obnovy by měl být brán zřetel již během návrhu a implementace celého řešení. Strategie musí zahrnovat kombinaci různých metod, aby bylo pokryto celé spektrum potenciálních problémů.

1. *Zálohování*

Pravidelné a časté zálohování je klíčové pro možnost obnovy jakéhokoliv IT systému. Pro činnost elektronického archívu to platí dvojnásobně a zálohování bylo zmiňováno již na řadě jiných míst tohoto dokumentu.
Je dobrá praxe uchovávat některé zálohy na jiném místě. Při výběru místa pro toto sekundární úložiště je nutné vzít do úvahy **polohu, dostupnost, bezpečnost, celkové prostředí** a **náklady**.

2. *Záložní řešení*

Přestože pravděpodobnost poruchy a nefunkčnosti celého systému nemusí být velká, je nutné s takovou alternativou v plánu pro mimořádné události počítat. Plán tedy musí vzít do úvahy provozování archívu na záložním řešení. V zásadě existují tři alternativy:

- Dedikované řešení vlastněné nebo provozované archívem
- Reciproční smlouva o pomoci a podpoře s jinou vnitřní nebo vnější jednotkou
- Smluvně pronajaté řešení pro takovéto mimořádné případy

Bez ohledu na to, pro kterou alternativu se archív rozhodne, musí být zařízení schopno zajistit funkčnost v rozsahu stanoveném v plánu pro mimořádné události.

3. *Náhrada (výměna) vybavení*

Pokud dojde k poškození nebo zničení IT vybavení anebo primární pracoviště je nefunkční či nedostupné, je nutné, aby náhradní hardware a software byl rychle získán (koupen), dodán a aktivován v příslušné lokalitě. Pro náhradu vybavení existují tři základní strategie. Při výběru vhodné strategie je nutné brát v úvahu, že možnosti dopravy mohou být silně omezeny v případě katastrofy.

- Servisní dohody s dodavateli
- Nákup a držení záložního řešení
- Dohody o využití existujícího a užívaného kompatibilního zařízení

4. *Úlohy a zodpovědnosti*

Koordinátor plánu pro mimořádné události musí při přípravě plánu určit týmy odpovídající za implementaci jednotlivých kroků strategie. Každý tým musí být řádně vyškolen a připravený k zapojení pro případ aktivace plánu. Pro efektivní nasazení musí členové týmu rozumět cílům týmu během obnovy, rozumět každému kroku, který je potřeba vykonat a znát úlohu svého týmu ve vztahu k činnosti ostatních týmů.

5. *Nákladové hledisko*

Při výběru vhodné strategie musí koordinátor plánu pro mimořádné události zvolit takové řešení, které lze efektivně implementovat s dostupnými lidskými a finančními zdroji. Během přípravy plánu pro mimořádné události by archív měl provést analýzu nákladů proti přínosům (cost-benefit analýzu) a vzít ji do úvahy při výběru optimální strategie obnovy.

2.4.5 Technologická infrastruktura

Při budování a provozování digitálního archivu vstupují technologie do hry na mnoha úrovních:

- Tvorba digitálního obsahu

- Konverze fyzického dokumentu na elektronický a vytváření asociovaných metadat
- Uložení elektronického dokumentu do archivu a manipulace s ním v rámci archívu
- Úkony spojené se správou elektronického dokumentu v archivu
- Vyhledávání a zobrazování uložených dokumentů
- Výstavba, provozování a udržování archivu