

5 Nákladový model

Vytvoření nákladového modelu digitálního archivu je složitá a komplexní úloha. Ke složitosti analýzy nákladů dlouhodobé digitální archivace přispívá řada faktorů. V prvé řadě je to nedostatek dat o nákladech na dlouhodobé archivování elektronických dokumentů. To nepochybně souvisí s tím, že koncept dlouhodobé archivace elektronických dokumentů je poměrně nový. Řada projektů je tudíž tak říkajíc v počátcích a tudíž nedisponuje spolehlivými čísly zachycujícími vývoj nákladů digitálního archivu v relevantně dlouhém období. Větší množství dat je dispozici pouze pro počáteční fáze archivního procesu - výběr dokumentů pro archivaci, konverze anebo změna formátu, přijetí dokumentu do archivu. Nicméně hlavní část nákladů je rozprostřena v dlouhodobém časovém horizontu a spočívá jednak v procesech zajišťujících dlouhodobé uchování správného bitové reprezentace daného elektronického dokumentu a jednak ve schopnosti tuto reprezentaci převést zpět do smysluplné informace. Dalším významným faktorem, který ovlivňuje budoucí vývoj nákladů jsou změny ve funkčnosti, výkonnosti a cenách informačních technologií, které jsou velmi dynamické a jejich dlouhodobý vývoj lze odhadovat pouze přibližně, pokud vůbec.

Při absenci empirických dat, lze budoucí náklady na digitální archivaci odhadnout, přičemž je nutné brát v potaz, jak nejasnosti a neurčitosti spjaté s digitální archivací přispívají k velké variabilitě těchto odhadů. Primárním zdrojem neurčitosti je samotný *raison d'être* pro digitální archivaci - stále se měnící a vyvíjející technologické prostředí v němž elektronické dokumenty existují. Tento fakt vyvolává trvalé nebezpečí, že archivované elektronické dokumenty se stanou nepoužitelné, protože hardware a/nebo software nutný k práci s nimi zastará anebo přímo přestane být funkční či dostupný. Samotná otázka technologického zastarávání je složitá. Minulá zkušenost ukazuje, že někdy období zastarávání může být překvapivě krátké, jindy naopak nečekaně dlouhé. Z právě řečeného je jasné, že apriorní odhad okamžiku, kdy současné hardwarové a softwarové prostředí bude muset být nahrazeno novým, je velmi široký a variabilní a tudíž dlouhodobý odhad výměn hardware a software a nákladů s tím spojených je velmi problematický a do každého nákladového modelu vnáší velký stupeň nahodilosti.

Druhým zdrojem nejistoty při odhadování nákladů na digitální archivaci je skutečnost, že tyto náklady jsou funkcí mnoha proměnných včetně doby archivace, archivačních technik, dále volby hardware a médií, úrovně přístupu, cílů archivace, včetně cílové klientské skupiny, konzervačních strategií, počtu a typů digitálních formátů, bohatosti metadat, pracovní náročnosti archivačních procesů, atd.. Je zřejmé, že je podstatně jednodušší vytvořit odhad nákladů pro jednu skupinu parametrů, než pro heterogenní model. Nicméně skutečnost je taková, že většina archivů, které by měly operovat v "reálném" prostředí je diverzifikovaných.

Nakonec, dalším komplikujícím faktorem při vytváření nákladového modelu digitální archivace je fakt, že máme zatím omezené zkušenosti s dlouhodobým uchováváním elektronických dokumentů. V současnosti předpokládáme, že existují tři hlavní konzervační strategie - zachování technologie, ve které byl předmětný elektronický dokument vytvořen, migrování formátu anebo emulace. Nicméně k pochopení všech

výhod a nevýhod jednotlivých konzervačních strategií a důsledků plynoucích z jejich přijetí je ještě potřeba udělat spoustu práce. Rozhodně nelze tvrdit, že přijetí a implementace kterékoliv z nich je dobře definovaný, zralý a dobře otestovaný postup. Tyto strategie, které by měly sloužit pro dlouhodobé uchovávání nikdy nebyly vyzkoušeny ani ve střednědobém produkčním prostředí. Není tudíž žádná záruka, že zítra se některá z nich neukáže být nevyhovující, zastaralou a nefunkční. Skutečnost, že velká část znalostí o digitální archivaci se stále vyvíjí a pravděpodobně dozná podstatných změn v následujících letech je dalším zdrojem neurčitosti v jakémkoliv nákladovém modelu pro dlouhodobou digitální archivaci.

5.1 Přehled nákladových indikátorů

V jádru každého digitálního archivu je dlouhodobá úschova na „bitové úrovni“, neboli v podstatě datové úložiště se svými vazbami na celkový systém správy archivu a dokumentů. Tato služba zahrnuje management dat a vytváření záloh, kontrolu integrity dat, pravidelné obnovování medií, obnovu po havárii, a podporu pro uložení a práci s metadaty. Datové úložiště nezajišťuje čitelnost dat ve smysluplné podobě, jinými slovy, jak je patrné z předchozích částí, nezabývá se problematikou migrace formátů a/nebo emulací a celou řadou dalších problémů. Tato omezení variability a zjednodušení umožňují nalézt práce publikující zkušenosti a čísla relevantní pro stanovení nákladů na nákladovosti provozování datového úložiště. Lze se zejména opřít o zkušenosti ze dvou pracovišť - Online Computer Library Center, Inc. (OCLC), Harvard University Library publikované Stephenem Chapmanem⁶⁸ a švédský Národní archiv ve Stockholmu publikované Jonasem Palmem⁶⁹. Vzhledem k tomu, že práce Jonase Palma ze švédského Národního archivu je novějšího data a navíc její výsledky v podstatě potvrzují a navíc i číselně korespondují s poznatky Stephena Chapmana, budeme se v dalších úvahách odvolávat hlavně na ni.

Pro ostatní části archivu existuje bohužel, jak již je řečeno výše, mnohem méně faktických podkladů a mnohem větší variabilita a tak naše odhady budou daleko volnější a spekulativnější. Často je obtížné přesně rozlišit mezi RMA (Records Management Application), lokálním úložištěm, digitálním archivem, důvěryhodným digitálním archivem, digitální úložištěm atd. Elektronické dokumenty během svého životního cyklu přecházejí mezi nimi a ne vždy a zcela jsou hranice zcela ostře vymezené. To je důsledkem i toho, že digitální uchování je nutností nejen pro archivní úschovu, ale i pro různorodé záměry ostatních podnikatelských subjektů.

Náklady zahrnuté v dlouhodobé archivaci elektronických dokumentů jsou ovlivněny řadou faktorů. Jejich shrnutí je uvedeno níže:

- Náklady na bezpečný digitální archiv a jeho funkcionality spojenou s dlouhodobým archivováním elektronických dokumentů
- Náklady na personální zajištění (lidské zdroje)
- Náklady na vývoj (nebo nákup) software a postupů na uchování elektronických dokumentů
- Náklady spojené s provozováním dlouhodobého digitálního archivu

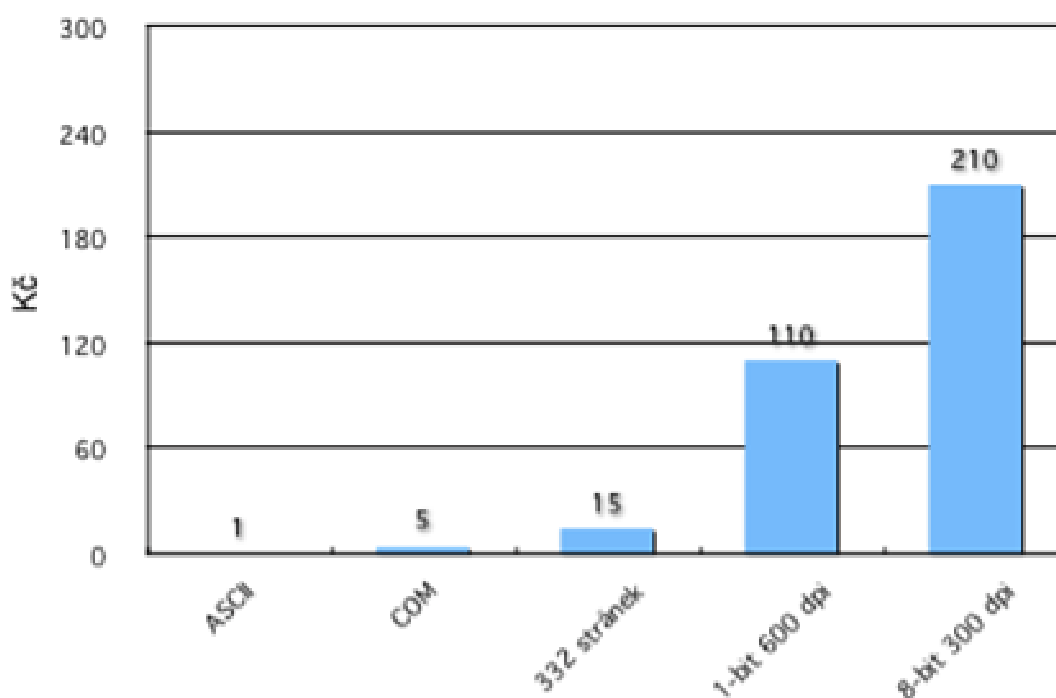
⁶⁸ Stephen Chapman, „Counting the Costs of Digital Preservation: Is Repository Storage Affordable?“ *Journal of Digital Information*, Vol. 4(2), May 2003

⁶⁹ Jonas Palm, „The Digital Black Hole“, 2006

- Další faktory ovlivňující celkové náklady

Při analýze nákladů spojených s archivací elektronických dokumentů je nutné rozlišovat dva typy elektronických dokumentů, kterými se archiv musí zabývat. Jednak ty, které přímo v digitální podobě byly vytvořeny a jsou proto v zásadě „malé“ a z hlediska nákladů na úschovu nepředstavují tudíž žádnou významnou položku, a pak dokumenty, které vznikly konverzí tradičních (zejména papírových) dokumentů. Konverzí papírového dokumentu na elektronický vzniká téměř vždy obrazový soubor, jehož velikost závisí na rozlišení, barevnosti, atd. nicméně jehož velikost je řádově několiknásobně větší než velikost dokumentu vznikajícího rovnou v digitální podobě a je tedy mnohem dražší na zpracování a uschování.

Ve švédském NA, podobně jako v OCLC, srovnávaly roční náklady na uložení 332 stránkové knihy. Následující graf ukazuje přibližné náklady spojené s jejím uchováním v různých formátech.



Čísla v grafu jsou pouze přibližná a vycházejí z obdobných údajů uvedených ve výše citovaných pracích. První sloupec ukazuje náklady na uložení 332 stránkové knihy, pokud by byla ukládána pouze jako ASCII text (3 KB/stránka). Druhý sloupec (COM) předpokládá digitalizaci knihy (scanováním), následné uložení nascanovaných obrázků na mikrofilm s využitím COM (Computer Output Microfilm) technologie a uchování mikrofilmů. Cena zde uvedená je pro speciální sklad na uchovávání filmů (regulovaná vlhkost). Pokud by byl mikrofilm uložen ve standardním skladu, byly by náklady přibližně poloviční. Třetí sloupec zachycuje náklady na uložení knihy tak, jak je ve standardním skladu, čtvrtý sloupec náklady na uložení v černobílém formátu a rozlišení 600 dpi. Poslední pátý sloupec pak předpokládá ukládání naskenovaných obrázků v 256 odstínech šedi a v rozlišení 300 dpi. Velikost ukládaného digitálního souboru a tudíž i náklady na uložení lze ovlivnit vhodnou volbou bezztrátového kompresního algoritmu, takže např. pro 1-bit scanovanou stránku v rozlišení 600 dpi lze dosáhnout velikosti 100 – 150 KB na stránku.

Pro úplnost dvě poznámky:

1. Vzhledem k tomu, že náklady na uložení se v podstatě lineárně odvíjejí od velikosti uchovávaného souboru byly by náklady na uložení 8-bit souboru v 600 dpi rozlišení čtyřikrát vyšší – tj. asi Kč 840.
2. Pokud bychom chtěli uložit barevný scan (obrázek) znamenalo by to pracovat se souborem 24-bit v rozlišení 300 dpi, tzn. se souborem třikrát větším a náklady na jeho uložení by tedy byly cca Kč 630.

Náklady na uložení zahrnují náklady systém na ukládání a správu dat, což zahrnuje kontrolu integrity dat, zálohovací procedury, automatické přenosy dat na nové pásky atd. Náklady na uložení dat v digitální formě (s výjimkou ASCII) jsou podstatně vyšší než je všeobecné očekávání. V roce 1996 Michael Lesk předpovídal že „... k protnutí nákladů na provozování digitální a tradiční knihovny dojde přibližně za pět let...“⁷⁰ a že digitální ukládání přinese „podstatnou cenovou výhodu“ během deseti let. Více jak deset let poté vidíme, že tato předpověď se splnila pro ASCII soubory, nicméně již pro 1-bit 600 dpi stránkové scany (cca 100 – 150 KB/stránka) to pravda není. To je způsobené skutečností, že v procesu uložení a správy dat je toho obsaženo více, než se obecně předpokládá. Obecná úvaha vychází z klesající ceny HW a zároveň se zvyšující se výkonnosti HW a permanentně narůstající kapacity úložných médií, jejichž kapacita se každé dva roky prakticky zdvojnásobuje. Zatímco na jedné straně cena HW vzhledem k výpočetní kapacitě stále klesá, množství dat, se kterými počítače musí pracovat stále narůstá a tudíž výpočetní kapacita potřebná pro zpracování souborů a dat se stále zvyšuje. Tento nárůst není způsoben pouze tím, že množství zpracovávaných informací se zvětšuje, ale též tím, že komplexnost jak reprezentace dat, tak systému a obslužných programů se vyvíjí a narůstá. Jonas Palm poznamenává, že hlavní část ceny systému úložišť dat nespočívá v úložných médiích, ale v SW a HW (a to nejen vlastní páskové knihovny a disková pole, ale také systémy určené pro správu úložiště, která se stává složitější právě s nárůstem uchovávaných dat).

Poznatky právě uvedené nastolují ihned několik otázek. Je vhodné naskenované dokumenty ukládat digitálně, eventuálně, které ano a které ne? Lze očekávat, že se v 10 - 20 letech náklady na uchovávání velkých, barevných scanů výrazně sníží? Není proto lepší pro takovéto dokumenty využít COM technologii, která též následně sníží náklady na konzervační postupy (migrace)?

5.2 Náklady na digitální archivní systém a jeho funkcionalitu spojenou s dlouhodobým uchováním elektronických dokumentů

Náklady na digitální archivní systém a jeho funkcionalitu spojenou s dlouhodobým archivováním tvoří řada různých komponent. Vytvořený nákladový model je nutně značně zjednodušený a celou řadu možností a detailů záměrně opomíjí, nicméně

⁷⁰ Michael Lesk, „Substituting Images for Books: The Economics for Libraries.“, presentation on Symposium on Document Analysis and Information Retrieval, April 15-17. 1996 at Las Vegas, Nevada

ukazuje váhu jednotlivých faktorů. Ukazuje též, které indikátory jsou ovlivněny tím, jak uchovávané elektronické dokumenty vznikly a též různými strategiemi uchování a které indikátory nejsou citlivé na tato rozhodnutí.

5.2.1 Fyzický prostor

- Serverovna, klimatizace
- Kancelářské prostory
- Konferenční místnost
- Příslušenství – záchody, kuchyň
- Zabezpečení

Fyzický prostor je potřeba pro umístění datového úložiště a systému dlouhodobé archivace. Servery jsou potřeba pro skladování elektronických dokumentů a pro řízení dlouhodobé archivace. Systémy pro vývoj a testování by měly být z bezpečnostních důvodů zcela odděleny od ostrého systému.

5.2.2 Hardware pro datové úložiště

- Disková pole
- Pásková knihovna
- Servery pro systémy správy úložiště a virtualizaci datového úložiště
- Disky, pásky a další úložná média
- Síť SAN a LAN

V závislosti na počtu elektronických dokumentů, celkové velikosti uchovávaných dat a předpokládaného ročního nárůstu je nutné zvolit vhodnou konfiguraci.

Švédský Národní archiv koupil pro zajištění archivace Hierarchical Storage Management System (HSM-system) před několika lety. Tento systém je postaven okolo „skladovacího robota“ – v tomto případě magnetopáskového systému propojeného s dalšími servery a počítači. Systém je postaven tak, aby detekoval a opravoval chyby v uložených digitálních informacích, mohl migrovat data na příští generaci systému hromadné úschovy, atd. Cena takového systému nespočívá v ukládacích médiích (tvoří 5%-10%), ale ve zbytku systému – hardware, software, podpoře, údržbě a administrativě.

Síť, její kapacita a konfigurace též závisí od typu a velikosti archivu – datové úložiště, které dostává velké množství digitálních objektů z mnoha různých míst bude potřebovat vysoko-rychlostní připojení, umožňující zvládnout rozdílné zatížení.

5.2.3 Software pro datové úložiště

- Operační systémy
- Zabezpečení
- Specifický software pro systém správy a virtualizaci datového úložiště

- Komunikační programy

Softwarovým základem pro datové úložiště je systém pro jeho správu. Kromě toho budou zapotřebí vhodné operační systémy a komunikační software. Datové úložiště bude nutné zabezpečit proti virům, neautorizovanému přístupu, změně integrity atd.

5.2.4 Hardware pro systém správy archivu a dokumentů

- Servery na přípravu a testování SW
- Servery pro systém správy dokumentů a digitálního archivu
- Pracovní stanice pro vývoj
- Pracovní stanice pro příjem dokumentů (včetně zařízení pro čtení podporovaných medií)
- Pracovní stanice pro správu archivu
- Diskové pole pro databáze
- Vybavení na vytváření záložních kopií
- Síťová komunikace

Systém uchování bude vyžadovat HW infrastrukturu, která jedna podpoří systémy správy dokumentů, jednak aplikace, které zajistí dlouhodobé uchování dokumentů a budou aplikovat tedy přijaté konzervační strategie (migrace, emulace atd.). Centrem systému správy dokumentů bude relační databáze, která bude obsahovat potřebné informace o uložených dokumentech. HW infrastruktura musí zajistit maximální spolehlivost, dostupnost a výkonnost této databáze. Systém uchování také bude potřebovat odpovídající lokální kapacitu pro práci s dokumenty.

Další servery a úložiště budou navíc potřeba v případě vývoje a testování. Pokud jsou předmětem vývoje a testování automatické nástroje je potřeba provádět testování na velkých objemech dat kvůli statické průkaznosti funkčnosti nástroje.

Systém uchování bude pravděpodobně muset umět číst páskové a diskové formáty, které budou dohodnuté pro fyzické předání dokumentů.

5.2.5 Software pro systém uchování

- Operační systémy
- Vývojářské prostředí
- Zabezpečení
- Staré a současné programové aplikace pro práci s elektronickými dokumenty
- Software na správu dokumentů a digitálního archivu
- Testovací prostředí a testovací software
- Komunikační programy
- Databáze

Dlouhodobé, spolehlivé a důvěryhodné uchování elektronických dokumentů může vyžadovat velmi komplexní programové vybavení. Nutností může být více operačních

systemů vzhledem k tomu, že systém archivu může pracovat v jiném operačním systému než v jakém byly některé elektronické dokumenty původně vytvořeny a/nebo uloženy a je nutné je převést (pro podporu různých prostředí může také archiv využívat prostředky pro vytvoření virtuálních systémů).

Systém uchování musí zkoumat různé způsoby uschování a experimentovat s řadou různých elektronických dokumentů a jejich formátů a i jejich dávek. Aby toto bylo možné bude se systém uschování muset starat o komplexní balík softwarových aplikací.

Vzhledem k tomu, že manuální zpráva velkého systému uchování je značně časově náročná a tudíž i nákladná, je její automatizace podstatným faktorem pro snižování nákladů zejména u rozsáhlých archivů. Automatizace konzervačních aktivit a automatické vyhodnocování aktivit může tudíž značně ovlivnit náklady spojené s digitálním uchováním.

V článku Jonase Palma je dále provedena detailní analýza všech nákladů provozování švédského digitálního národního archivu na příštích pět let. Roční nárůst úložné kapacity se počítá 40 Tb za rok.

Specification of costs	1st year	2nd year	3rd year	4th year	5th year	5 years
1 HSM storage system						
price 2003 + 3% interest/year						
€406,643 spread over 5 years, 81,328/year	94818	92379	89939	87499	85059	449694
Staff for operation, 0.6 fte, €40,000/fte incl all costs	24000	25200	26400	27660	28980	132240
Premises 100m ² , €126 per m ²	12600	12915	13237	13568	13908	66228
Service/support	22700	28900	28900	28900	28900	138300
Total storage costs	154118	159394	158476	157627	156847	786462
Annual storage cost per Gb	3.85	1.99	1.32	0.98	0.78	
Average storage cost per Gb for 5 years						7.86
Storage medium 40 Tb/yr	17930	11295	7116	4483	2824	43648
Staff for input, 0.4 fte €40,000/fte incl all costs	16000	16800	17600	18440	19320	88160
Yearly Input Cost (staff, storage medium)	33930	28095	24716	22923	22144	131808
Cost of input per Gb	0.84	0.7	0.61	0.57	0.55	0.66
Total cost per newly added Gb	4.69	2.69	1.93	1.55	1.33	
Average total cost per GB for 5 years						9.18

Podívejme se na některé závěry vyplývající z této analýzy:

- Průměrná cena uložení 1 GB po dobu 5 let je € 9,18.
- V celkových ročních nákladech tvoří náklady na systém správy úložiště 49% a náklady na lidské zdroje (operátoři, vkládání dat a údržba) 39%. (V našich podmínkách by v danou chvíli a pro nejbližších 5 let toto procento bylo pravděpodobně asi poloviční – švédský model vychází z předpokladu, že ho jeden člověk celkově přijde na 40.000 € za rok.)
- Vyčíslené náklady vychází z malého stupně aktivity archivu směrem ven, tj. k uživatelům. Pokud by uložené informace začaly být více využívány, vzrostly by náklady na administrativu i náklady na další externí servery na nichž by informace byly zpřístupňovány. S narůstajícím počtem archivovaných informací budou nicméně aktivity směrem ven v absolutní hodnotě narůstat, i když v relativní hodnotě může být malá aktivita zachována.

Tímto se dostáváme k další podstatné součásti nákladů a to jsou náklady na lidské zdroje.

5.3 Osobní náklady

V této části bude provedena analýza povinností zaměstnanců archivu a její implikace na náklady. Bude proveden rozbor potřebných kvalifikací a počtů zaměstnanců. Příložený nákladový model pak analyzuje časové požadavky na jednotlivé zaměstnance v závislosti na jejich kvalifikaci.

5.3.1 Pracovní úkoly zaměstnanců důvěryhodného digitálního archivu

- Sběr požadavků
- Získávání financování a podpory
- Návrh a vybudování digitálního archivu a datového úložiště

Personál bude muset začít s návrhem a budováním digitálního archivu. To bude vyžadovat rozpočet v odhadu asi 2 - 3 člověkoroků. I návrh archivu koupený od třetí strany bude vyžadovat modifikace, aby byly splněny specifické požadavky pro danou organizaci.

- Nastavení procesů
- Průběžné řízení a provozování digitálního archivu a datového úložiště
- Průběžné řízení bezpečnosti
- Vývoj systému pro kontrolu kvality a sepsání provozní a řídicí dokumentace
- Vytvoření standardních operačních postupů
- Uživatelské manuály

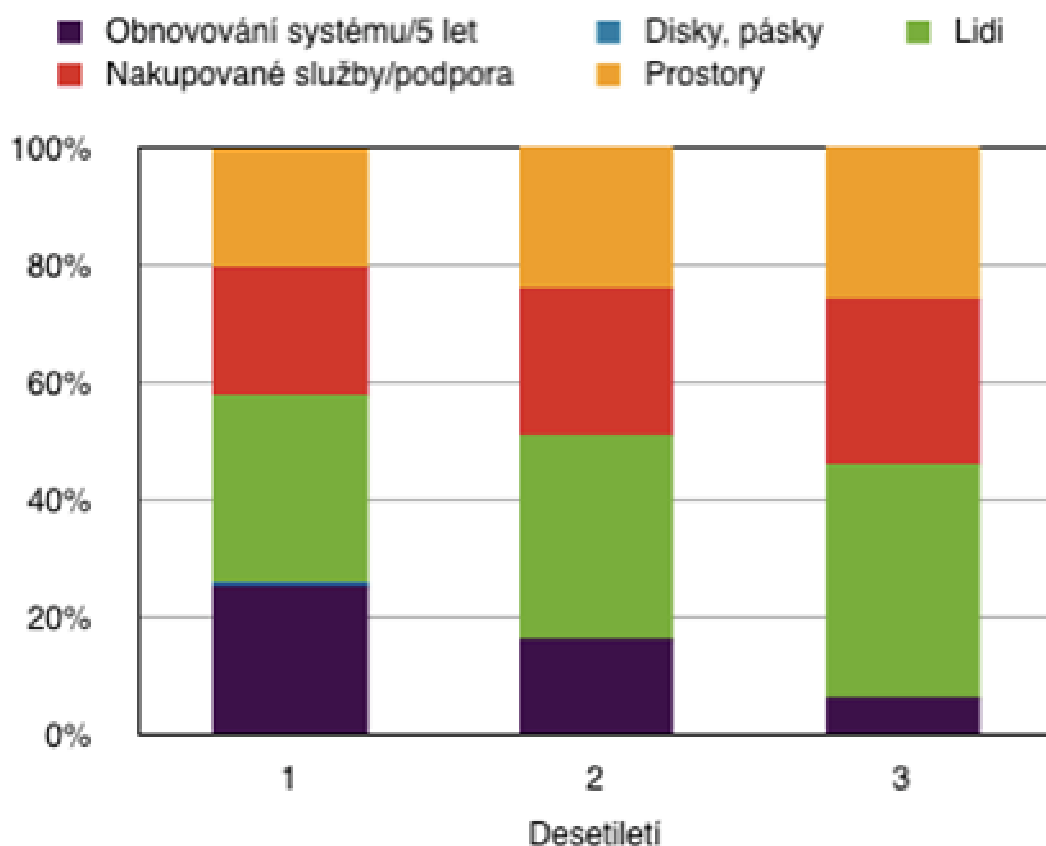
Jakmile je digitální archiv postaven, je dalším krokem vytvoření relevantních procedur a uvedení archivu do provozu. Interní management vytvoří a implementuje bezpečnostní a interní procesy, komplexní systém řízení kvality, každodenní Standardní Operační Postupy a uživatelské manuály. Nakonec musí začít vyvíjet postupy pro uchování, vyhodnocovací testy a začít implementovat postupy pro udržitelné uchování dokumentů.

5.3.2 Povinnost a úkoly zaměstnanců služeb pro zákazníky

- Příjem a zpracování dokumentů
- Správa přístupu k uchovávaným dokumentům
- Školení a výcvik

Jak už zmiňujeme v předchozím odstavci pro švédský NA vychází v danou chvíli náklady na lidské zdroje na 39% celkových nákladů. Pro další roky počítají s nárůstem tohoto podílu, který bude způsoben jednak růstem mezd a jednak nárůstem objemu a počtu aktivit a vzhledem k tomu nárůstem počtu pracovníků.

Švédský národní archiv se pokusil i o odhad poměru nákladů v příštích 30 letech. Jak je z následujícího grafu patrné, poměrné náklady na obnovu systému a úložná média budou klesat, poměrné náklady na fyzické prostory mírně vzrostou a podstatně vzrostou náklady na nakupované služby a lidské zdroje.



5.4 Náklady na vývoj a/nebo nákup software a postupů na uchování elektronických dokumentů

- Určení požadavků autenticity
- Analýza požadavků na autentičnost

V mnoha případech lze vlastnosti elektronických dokumentů podstatně pro integritu a autenticitu dokumentu oddělit od ostatních, méně důležitých vlastností. Digitální konzervační aktivity lze pak soustředit na vlastnosti klíčové pro integritu a autenticitu dokumentu.

Výběr klíčových vlastností pro integritu a autenticitu dokumentů může úspěšně provést pouze organizace, která dokumenty vytvořila. Náklady spojené s určením autenticity záznamů se objevují v nákladech na přijetí dokumentu do archivu.

V příloženém nákladovém modelu se předpokládá, že autentičnost je nutné určit pro každou novou dávku přijímaných dokumentů. (Dávka obsahuje dokumenty vytvořené stejnou aplikací, jejichž přijetí probíhá na jednom místě a ve stejnou dobu.) V nákladovém modelu je ukázáno, že velikost dávky má podstatný vliv na celkové provozní náklady.

- 1) Návrh postupu uchování
- 2) Vývoj postupu uchování

3) Uchovávací software

4) Prohlížeč/renderovací software

Po té co otázka autenticity byla rozhodnuta je dalším krokem návrh a posléze vývoj vhodného konzervačního postupu. Do budoucna lze očekávat, že budou vyvinuta komerční portfolia různých konzervačních strategií. Nicméně i potom bude nutné vyhodnotit vhodnost různých strategií vzhledem ke specifické dávce dokumentů přijímaných do archivu a v některých případech bude nutné postup značně upravit anebo zcela změnit.

5) Otestování postupu

6) Pokud je schválen, pokračování

7) Pokud není schválen, návrat k bodům 1, 2 nebo 3

8) Dokumentování postupu

Každý postup nebo zvolená strategie musí být otestovány a zdokumentovány. Všechny IT operace prováděné v rámci jakéhokoliv bezpečného digitálního archivu musí odpovídat nejpřísnějším standardům kvality. Pro zachování autenticity je vysoká úroveň kvality nezbytnou podmínkou, jelikož systém kvality a dokumentace slouží k prokázání toho, že konzervační úkony dosáhly zamýšlených výsledků a že nijak neovlivnily ostatní dokumenty.

5.5 Náklady spojené s provozem archivu

V tomto odstavci se budeme věnovat jednak nákladům na digitalizaci a v druhé části se zaměříme na náklady spojené s prováděním konzervačních/archivačních operací na elektronických dokumentech.

5.5.1 Digitalizace

Proces digitalizace se týká dokumentů, které vznikly v jiné než digitální podobě. Nejčastěji se jedná o papírové dokumenty, ale může se též jednat o analogové audiovizuální záznamy. Proces digitalizace pokrývá řadu aktivit - výběr dokumentu, tvorba popisu a metadat a samozřejmě skutečnou konverzi do digitálního formátu. Při konverzi závisí výsledná kvalita, ale též velikost výsledného digitálního záznamu, na použitém zařízení, konvertovaném dokumentu, procesních specifikacích, atd. Mnohdy (obraz, zvuk) je pro výběr vhodného zařízení, procesů, nastavení kontroly kvality nutné expertní posouzení obsahu a kvality konvertovaného dokumentu. V obvyklých případech bude toto posouzení pravděpodobně zajišťovat archiv pomocí vlastních pracovníků, v méně obvyklých, bude muset pravděpodobně použít externích služeb.

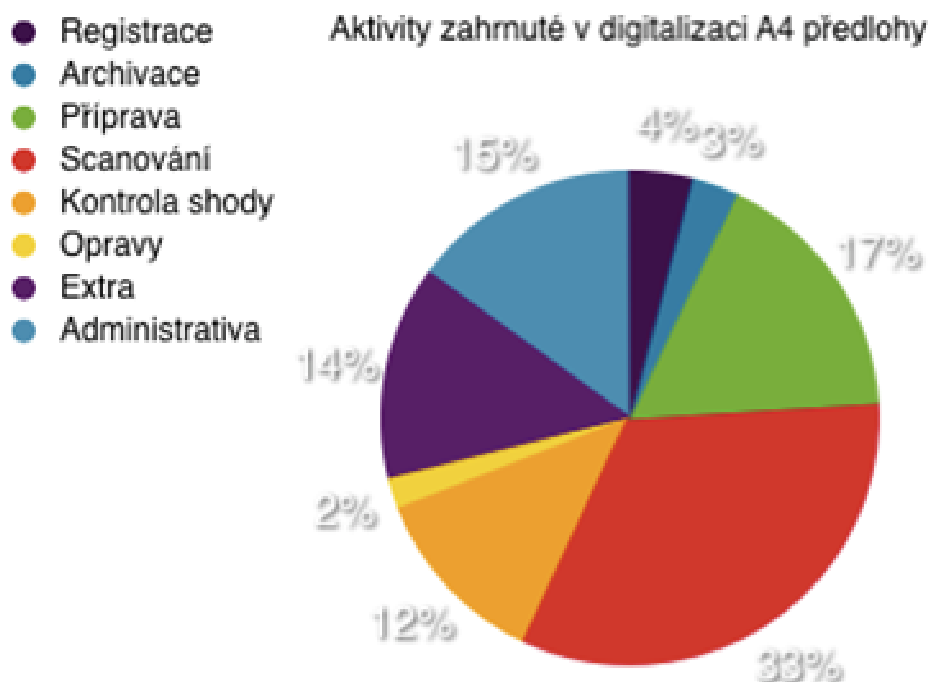
V dalším se zaměříme na papírové dokumenty a jejich scanování a opřeme se přitom opět o data poskytnuté švédským NA. Pro ilustraci se podívejme na dva příklady, které stojí na opačných koncích možného spektra:

a) A4 předlohy scanované jako 1-bitové obrazy v rozlišení 600 dpi

b) A2 předlohy scanované jako 8-bitové obrazy (tzn. v 256 úrovních šedi) a v rozlišení 297 dpi.

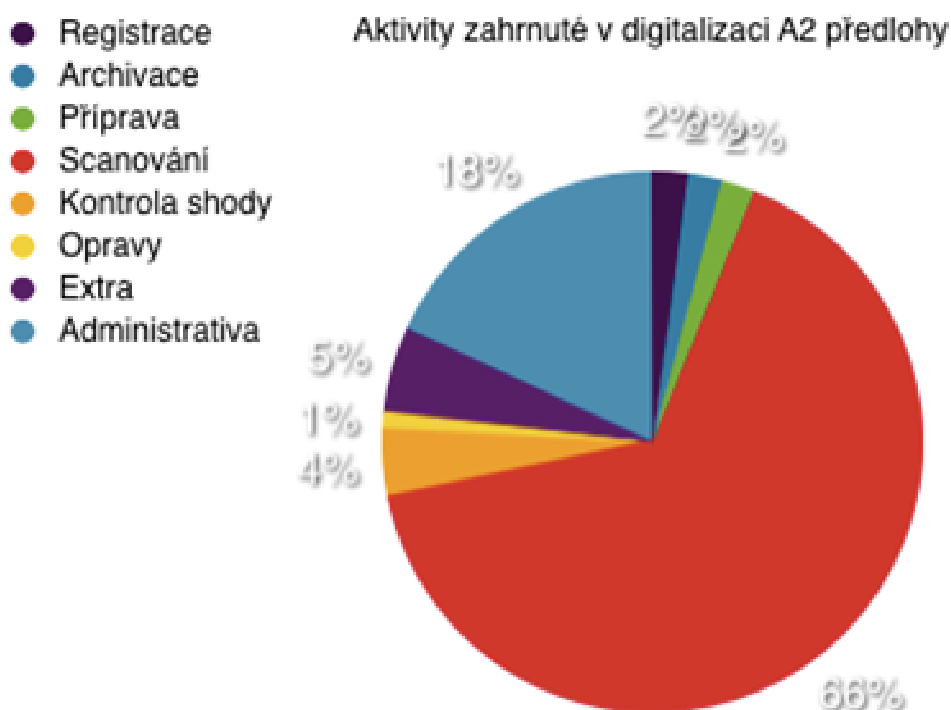
Ad a) V švédském NA scanují ročně cca 5 milionů A4 předloh v této kvalitě. Proces probíhá na automatických scannerech, tudíž v podstatě automaticky a náklady na

digitalizaci jedné stránky jsou asi 0,10 Euro. Následující graf ukazuje poměrnou distribuci nákladů.



Jak je vidět, příprava, kontrola shody, extra práce a administrativa jsou přibližně stejně zastoupeny a dohromady reprezentují více jak polovinu nákladů. Vlastní scanování reprezentuje přesně třetinu nákladů.

Ad b) Předlohy o velikosti A2 jsou většinou mapy, kresby, plány, apod. a pro jejich scanování se používají neautomatické scannery. Počet takto zpracovaných je cca 1,3 milionů předloh za rok a náklady na jednu předlohu jsou cca 0,61 Euro. Jak ukazuje následující graf, distribuce nákladů se podstatně liší od předchozího případu.



Jak je vidět, je distribuce nákladů zcela jiná. Vlastní scanování tvoří dvojnásobek než v předchozím případě, jasně druhé místo zaujímá administrativa s 18%, ostatní náklady se pohybují na úrovni 1% - 5%.

5.5.2 Konzervační a archivační aktivity

V této části se zaměříme na náklady spojené s prováděním konzervačních a archivačních operací na elektronických dokumentech. Půjde zejména o tyto aktivity:

- Migrace dokumentu
- Migrace dokumentu na žádost
- Emulace

Rozdíly mezi těmito aktivitami vysvětlíme s pomocí OAIS terminologie. Migrace nebo jiná transformace elektronického dokumentu vede ke změně archivního balíku (AIP) uloženého v archivu. AIP1 se změní na AIP2 a AIP2 slouží jako základ pro předávacího balíku (DIP), který je vydáván uživateli archivu.

“Migrace na žádost” nemá na AIP žádný vliv, ovlivňuje pouze výsledný vydaný DIP podle potřeb, postavení a oprávnění klienta. Je samozřejmě nezbytné zajistit, aby vydaný DIP byl autentický a smyslově přístupný. Pokud klient si vyžádá přístup k určitému dokumentu, management archivu vybere vhodný softwarový nástroj, provede transformaci dokumentu a klientovi je zpřístupněn výsledek této transformace.

Získání přístupu k elektronickému dokumentu prostřednictvím emulace nemá žádný vliv na elektronický dokument obsažený v AIP a v principu ani DIP se v čase nemění. Archiv (nebo třetí strana) ale musí poskytnout uživatelům vhodné SW prostředky (tj. emulátor), které umožní daný dokument zobrazit a případně další operace s ním. Jak ukazuje nákladový model, náklady na výše zmíněné aktivity tvoří ne příliš podstatnou část celkových nákladů digitálního bezpečného archivu.

- 1) Identifikace elektronických dokumentů k archivačním operacím
- 2) Obdržení archivního balíku obsahujícího daný dokument
- 3) Výběr konzervační strategie a postupu

Elektronické dokumenty, které potřebují transformaci anebo pro které musí být vyvinut emulátor může archivní systém vybírat automaticky anebo mohou být identifikovány ručně. Archivní balíky obsahující dokumenty potřebující nějaký zásah musí být nejprve vybrány a převedeny do dočasného úložiště. Následně lze pro ně zvolit vhodnou konzervační strategii.

- 4) Příprava na transformaci
- 5) Přidání metadat
- 6) Oprava nebo modifikace elektronického dokumentu
- 7) Transformace elektronického dokumentu s použitím zvolené metody
- 8) Vyhodnocení výsledků transformace
- 9) Elektronický dokument je přístupný a smysluplně čitelný s pomocí vhodného software
- 10) Autenticita elektronického dokumentu byla zachována

- 11) Pokud NE, návrat do bodu 6 anebo k ještě dřívějším krokům
- 12) Aktualizace archivního balíku výsledným elektronickým dokumentem a uložení zpět do datového úložiště

Každá transformace musí být vyhodnocena a musí být prokázáno, že transformované elektronické dokumenty jsou nadále přístupné a dále též, že jejich autenticita a integrita nebyla narušena.

Po úspěšné transformaci jsou transformované elektronické dokumenty aktualizovány v archivním balíku a tento pak vložen zpět do datového úložiště.

5.6 Další faktory, které ovlivňují celkové náklady

Kromě faktorů rozebíraných v předchozích odstavcích, které mají v podstatě přímý vliv na náklady, existuje ještě řada faktorů, které mají vliv nepřímý, nicméně celkové náklady mohou podstatně ovlivnit.

5.6.1 Veřejné služby

- Počet uživatelů digitálního archivu
- Hardware a software zajišťující klientský přístup k záznamům
- Potřebná podpora a údržba
- Školení

Počty uživatelů a rozsah služeb jim poskytovaných podstatně ovlivní celkové náklady archivu. Na druhé straně, pokud by tyto služby bylo možné zpoplatnit anebo přímo poskytovat na komerční bázi, bylo by možné část nákladů takto kompenzovat.

5.6.2 Frekvence konzervačních aktivit

Na jedné straně časté provádění konzervačních akcí nad skupinami elektronických dokumentů vede k nárůstu nákladů, na straně druhé, příliš dlouhé intervaly mezi konzervačními zásahy zvyšuje rizika spojená s autenticitou, integritou a čitelností uchovávaných dokumentů a potažmo vytváří riziko dalších neplánovaných nákladů.

5.6.3 Monitorování technologického pokroku

S předchozím odstavcem souvisí sledování změn hardware, software a dalších systémů archivu, s ohledem na jejich zastarávání a potenciální rizika pro dostupnost archivovaných elektronických dokumentů. Součástí těchto aktivit je návrh vhodných konzervačních aktivit pro ohrožené elektronické dokumenty a jejich implementace.

5.6.4 Dodatečné úložné kapacity

Praxe ukazuje, že kromě vytváření záložních kopií aktuálního stavu digitálního archivu, většina organizací udržuje i všechny průběžné kopie elektronického dokumentu v průběhu jeho životního cyklu. Tzn. archiv stále uchovává elektronický dokument ve všech formách a formátech, ve kterých se elektronický dokument v průběhu životního cyklu nacházel a to včetně forem analogových (např. papírových u dokumentů), ve kterých byl před digitalizací. To samozřejmě vede k několika

násobnému navýšení potřebné úložné kapacity a též, i když ne tak výraznému, nárůstu archivních aktivit.

5.7 Nákladový model

V 6.7 je základní číselný model nákladů spojených s vybudováním a provozem digitálního archivu. Model předpokládá vybudování archivu, který je schopen archivovat emaily, textové dokumenty, spreadsheets, databáze a skeny papírových dokumentů. Kapacita ani vybavení archivu nepředpokládá ukládání digitalizovaných audio-vizuálních objektů. Model se též nezabývá s procesy spojenými s digitalizací papírových dokumentů a uchováváním různých obrazových formátů, i když jeho úložná kapacita je navržena tak, aby byl schopen přijímat každý rok několik milionů dokumentů naskenovaných v různých formátech a samozřejmě je též dlouhodobě uchovávat.

Jedním z cílů modelu bylo podívat se na závislost časových nároků a nákladů na typu dokumentu a též na formátu, v kterém je archivem získán na formátu do kterého je dokument migrován. V zásadě se uvažují situace, kdy dokument je získán v nativním formátu příslušného programu (tzn. např. spreadsheet vytvořený v Excelu, je vkládán do archivu jako XLS soubor), a dokument se migruje do nových nativních formátů, tak jak je příslušný software inovován. Druhá možnost je, že soubor je při příjmu do archivu migrován do XML formátu a je uchovávan a migrován jako XML soubor. Další alternativa je, že soubor je již při příjmu v XML formátu. Jak model ukazuje náklady i pracnost se pro různé formáty značně odlišují. Hlavním důvodem této odlišnosti jsou různé časové intervaly mezi migracemi pro jednotlivé formáty – 3 roky pro nativní formáty, 10 let pro XML formát. Je dobré si uvědomit, že při tříletém cyklu pro migraci, po deseti letech činnosti archivu migruji v jedenáctém roce již tři ročníky souborů uchovávaných v nativních formátech a pouze jeden rok souborů uložených v XML formátu, po dvaceti letech je to šest ročníků souborů v nativních formátech a stále pouze jeden v XML formátu.

Model se skládá ze třech archů. První arch shrnuje předpoklady ohledně hardwarového a softwarového vybavení a personální a prostorového zabezpečení. V druhém archu jsou udělány určité předpoklady ohledně typu a formátů archivovaných dokumentů a intervalů migrace a pracnosti určitých úkonů a na základě toho jsou vypočteny časové nároky související s příjmem dokumentů do archivu, s údržbou archivu po dobu 20 let v závislosti na typu a formátu dokumentu a časové nároky na vývoj nových postupů konzervace, opět v závislosti na typu a formátu dokumentu. Třetí sheet vychází z výpočtů časové náročnosti a je v něm vyčíslena náročnost finanční.