

Příloha č.3

ČÁST III. - DIGITALIZACE A UKLÁDÁNÍ DAT

Verze dokumentu

Následující tabulka obsahuje informace o verzi dokumentu:

| Verze | Datum | Popis změny v dokumentu |
|-------|-------------|--|
| 0.1 | 18. 7. 2010 | První draft |
| 1.0 | 30. 7. 2010 | Druhý draft |
| 1.2 | 3. 8. 2010 | Dokument k připomínkování -technické kapitoly |
| 1.4 | 8. 8. 2010 | Doplnění obecných kapitol |
| 1.5 | 10. 8. 2010 | Doplnění nových kapitol |
| 1.6 | 10. 8. 2010 | Doplnění nových kapitol |
| 1.7 | 11.8.2010 | Doplnění nových kapitol |
| 1.8 | 15.8.2010 | Doplnění nových kapitol, doplnění a úprava kapitoly 4 a 7 |
| 1.9 | 15.8.2010 | Doplnění nových kapitol (3.9, 7.2.3), doplnění/úprava kapitoly 4.1.1.1.7 |
| 1.11 | 17.8.2010 | Příprava dokumentu k připomínkování |
| 3.0 | 20.8.2010 | Změna struktury dokumentu, doplnění |
| 3.1 | 25.8.2010 | Finalizace, verze pro akceptaci |
| 3.2 | 28.8.2010 | Formální úpravy |
| 4.0 | 4.9.2010 | Přeřazení kapitoly z hlavního dokumentu |
| 4.1 | 8.9.2010 | Úpravy hodnocení variant |

Tabulka 1: Verze dokumentu

Obsah

| | | |
|-------|--|----|
| 1 | Technické řešení..... | 5 |
| 1.1 | Vlastní koncepce řešení..... | 5 |
| 1.1.1 | Návrh a popis architektury řešení..... | 14 |
| 1.1.2 | Variantní návrhy technického řešení – HW/SW/data..... | 55 |
| 1.1.3 | Naplnění požadavků typizovaného projektu..... | 56 |
| 1.2 | Porovnání variant technologických řešení..... | 56 |
| 1.2.1 | Srovnání nabídek jednotlivých dodavatelů..... | 56 |
| 1.2.2 | Výhody a nevýhody jednotlivých řešení..... | 62 |
| 1.2.3 | Analýza technických a bezpečnostních rizik..... | 62 |
| 1.3 | Doporučení a upřesnění pro účely zadávací dokumentace a realizační projektové dokumentace..... | 62 |
| 1.3.1 | Specifikace zadání technického řešení..... | 62 |
| 1.3.2 | Požadavky na implementaci, školení a technickou podporu..... | 63 |
| 1.4 | Provozní zajištění projektu..... | 63 |
| 1.4.1 | Potřebné energetické a materiálové toky..... | 63 |
| 1.4.2 | Záruky a servis..... | 64 |
| 1.4.3 | Údržba a nákladnost oprav..... | 64 |
| 1.4.4 | Údaje o životnostech jednotlivých zařízení..... | 64 |
| 1.4.5 | Údaje o provozním zajištění SW a datových komponent..... | 64 |
| 1.4.6 | Změny v provozní náročnosti vlivem opotřebení..... | 64 |

Seznam tabulek

| | | |
|-------------|--------------------------------------|----|
| Tabulka 1: | Verze dokumentu..... | 2 |
| Tabulka 2: | Výstup digitalizačního procesu..... | 8 |
| Tabulka 3: | Dokumenty k digitalizaci..... | 15 |
| Tabulka 4: | Digitalizační metriky knih..... | 17 |
| Tabulka 5: | Digitalizační metriky periodik..... | 18 |
| Tabulka 6: | Roční přírůstek do KDS..... | 23 |
| Tabulka 7: | přírůstky a úbytky KDS..... | 23 |
| Tabulka 8: | Varianty kvality digitalizace..... | 26 |
| Tabulka 9: | Objemy zdravotní dokumentace..... | 28 |
| Tabulka 10: | Objem dat pro KDÚ za 5 letročně..... | 28 |

| | |
|--|----|
| Tabulka 3: Rozdíly mezi jednotlivými subsystémy..... | 42 |
| Tabulka 4: Přehled nabídek pro KDS..... | 59 |
| Tabulka 5: Přehled nabídek pro KDR..... | 60 |
| Tabulka 6: Ceny za digitalizaci dokumentů..... | 60 |
| Tabulka 7: Ceny za KDJ..... | 61 |

Seznam obrázků

| | |
|---|----|
| Obrázek 1:Koncept České digitální knihovny..... | 6 |
| Obrázek 2: Metodika digitalizace | 7 |
| Obrázek 3: Propojení ESS s KDS a NDA..... | 12 |
| Obrázek 4: Architektura KDS a KDR..... | 34 |
| Obrázek 5: Architektura KDÚ..... | 35 |
| Obrázek 6: Funkční model OAIS..... | 36 |
| Obrázek 7: Technologické schéma ukládání..... | 48 |
| Obrázek 8: Technologické schéma mobilní KDJ | 50 |

1 Technické řešení

V rámci této kapitoly je popsáno technické řešení projektu. Struktura popisu technického řešení je dána přílohou příručky pro žadatele č. P1b Osnova studie proveditelnosti. Součástí této kapitoly je několik variant technického řešení tohoto projektu.

1.1 Vlastní koncepce řešení

Vymezení důležitých pojmů

Systém digitální archivace - Digitální archivace pokrývá identifikaci, ohodnocení, popis a označení, uložení, uchování, řízení a získávání digitálních dokumentů a zahrnuje veškeré politiky, návody, průvodce a systémy asociované s těmito procesy. Trvale (dlouhodobě) zajišťuje nejen fyzickou, ale i logickou dostupnost a čitelnost dokumentů.

Systém dlouhodobého uložení - Systém dlouhodobého uložení zajišťuje, že uložená data v podobě sekvence bitů zůstanou po dlouhou dobu přístupná a neměnná. Systém dlouhodobého uložení je základní nutnou komponentou digitální archivace. Sám o sobě však nemůže zajistit dlouhodobou dostupnost a plné pochopení digitálních dokumentů v budoucnosti.

Krajské negarantované úložiště - Realizováno v rámci TC kraje jako pomocné úložiště aplikací provozovaných v rámci veřejné správy na území kraje. Má sloužit pro ukládání bezpečnostních kopií dat těchto aplikací. Není určeno pro dlouhodobé uložení dat. Toto úložiště bude využívat i systém krajského digitálního úložiště KDÚ.

Data v negarantovaném úložišti jsou přístupná pouze prostřednictvím aplikací, které si sem tato data uložily.

Krajské garantované archivní úložiště - Realizováno v rámci TC kraje jako společné dlouhodobé bezpečné úložiště pro dva archivní systémy vybudované podle modelu OAIS:

- **Krajskou Digitální Spisovnu (KDS)**
- **Krajský Digitální Repozitář (KDR)**

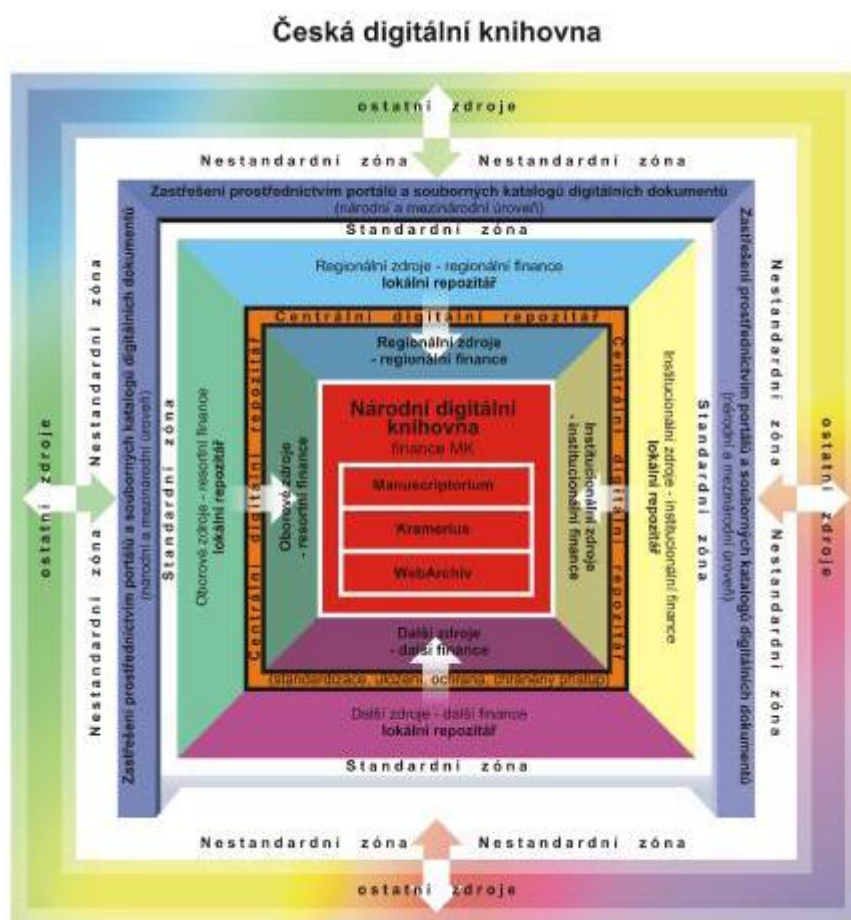
Krajské digitální úložiště (KDÚ) - Realizováno v rámci TC kraje jako dlouhodobé bezpečné úložiště dat pocházejících z nejrůznějších informačních systémů, kde budou uložena data nejrůznějšího typu především z toho důvodu, aby se včas zabránilo jejich zničení před definitivním rozhodnutím, jak s nimi bude naloženo později. Hlavním úkolem KDÚ je dlouhodobé bezpečné uložení počítačových dat spočívající v zabezpečení neměnnosti uložených dat ve formátu, v jakém byla uložena.

Problematika digitalizace

V problematice digitalizace se na úrovni kraje v současné době prolínají minimálně dva projekty:

1. Projekt vytvoření Národní digitální knihovny,
2. Projekt vybudování Krajského technologického centra a jeho části "Digitalizace a ukládání" (předmět této studie).

Oba dva projekty mají být součástí širší koncepce s názvem Česká digitální knihovna, která je tvořena velkým množstvím dalších digitálních dokumentů oborového, regionálního, institucionálního i jiného charakteru. Pouze některé z těchto zdrojů se kvalifikují jako nejcennější součást - jádro národního kulturního dědictví umístěné v Národní digitální knihovně (na následujícím obrázku uprostřed). Za jeho vytvoření a uchování nese odpovědnost Národní knihovna a Ministerstvo kultury. Za shromažďování, trvalé uchovávání (na centrálním datovém úložišti nebo lokálních datových úložištích) i zpřístupnění zdrojů mimo „jádro“ národního kulturního dědictví nesou odpovědnost resortně příslušná ministerstva, regiony, instituce atd. Schematicky je koncept České digitální knihovny naznačen na následujícím obrázku.



Obrázek 1:Koncept České digitální knihovny

Oba projekty by měly být koordinovány tak, aby se pokud možno nepřekrývaly ani v části digitalizace ani v části uložení. Pro koordinaci digitalizace knihovných fondů jsou připravovány dva nástroje: registr digitalizace a souborný katalog. Paralelně je řešena problematika jednoznačných identifikátorů digitalizovaných objektů.

Projekt Národní digitální knihovny je tvořen vybraným souborem publikovaných digitálních i digitalizovaných dokumentů kvalifikovaných jako základ – jádro národního kulturního dědictví určený k trvalému uchování a zpřístupnění současným i budoucím uživatelům.

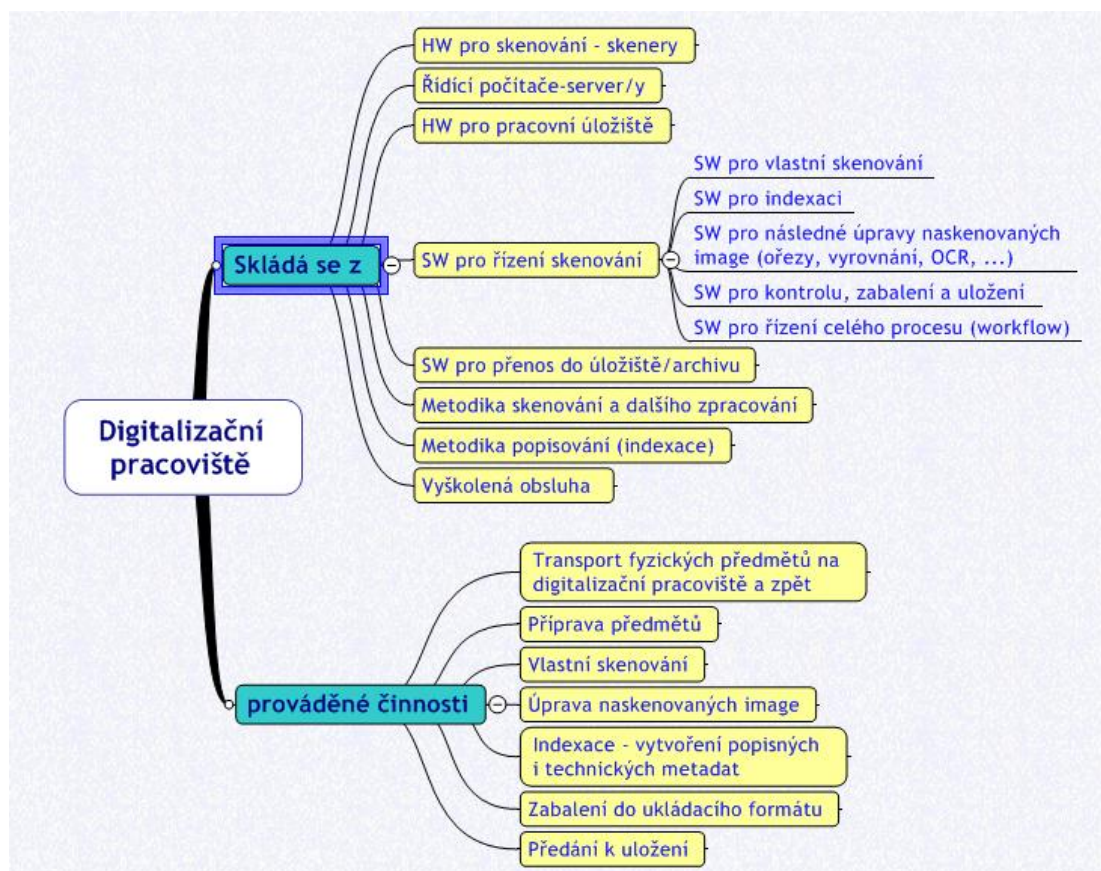
Vytvoření Národní digitální knihovny se zabývá digitalizací, uložením a zpřístupněním podkladů celonárodního významu v následujícím členění:

- Digitalizace historických bohemikálních dokumentů v úplnosti
- Digitalizace bohemikálních dokumentů 19. stol. v úplnosti
- Digitalizace soudobých přírůstků bohemikální literatury
- Digitalizace bohemikálních dokumentů 20. stol. výběrově

Projekt zahrnuje podle odhadu zpracování cca 1,2 milionů dokumentů (svazků) což představuje cca 350 milionů stran.

Metodika digitalizace knih a listinných památek

Problematika vlastní digitalizace knih a listinných památek je připravována, ověřována aktivitami Národní knihovny, Moravské zemské knihovny i Krajských knihoven. Několik národních grantových projektů umožnilo odstartovat projekty digitalizace v knihovnách ČR již počátkem 90. let minulého století. Od samého počátku byly respektovány mezinárodní standardy a díky tomu je možné všechny výstupy integrovat do různých národních i nadnárodních portálů. Metodika digitalizace se průběžně upravuje podle získaných zkušeností. Metodika zahrnuje jak zřízení digitalizačního pracoviště (výběr skenerů, další potřebný hardware, software pro řízení, metodika zpracování naskenovaných předloh), tak jeho provoz a procesy, které ho podporují.



Obrázek 2: Metodika digitalizace

Metodika je připravována jak pro digitalizaci knih, tak pro digitalizaci jiných předloh (např. map). Výstupem z digitalizačního procesu je ke každému svazku několik souborů podle následující tabulky:

| | Typ souboru | | Formát |
|---------|------------------------|---|----------|
| MC | Master Copy | Archivní kopie | JP2 |
| UC | User Copy | Uživatelská kopie pro systém zpřístupnění | PDF HT |
| OCR | TextFormat | Textový soubor | ALTO XML |
| DMD XML | DescriptiveMetadata | Popisná metadata | XML |
| AMD XML | AdministrativeMetadata | Administrativní metadata | XML |

Tabulka 2: Výstup digitalizačního procesu

Z pohledu ukládání, resp. dlouhodobé archivace tyto soubory tvoří jeden digitální objekt („informační balíček“ podle modelu OAIS), který má být uložen v (dlouhodobém) archivu.

Na úrovni Kraje bude existovat jedno digitalizační pracoviště, které bude předávat výstupy své činnosti k archivaci do Krajského digitálního repozitáře KDR. Všechny výstupy by měly být podchyceny v Registru Digitalizace provozovaném Národní knihovnou.

Pro projekty digitalizace na Kraji doporučujeme převzít metodiku digitalizačních pracovišť zpracovanou Národní knihovnou, včetně standardů pro výstupní datové formáty.

Problematika ukládání a archivace

Díky nárůstu využívání moderních informačních technologií a stále častějšímu používání elektronických forem dokumentů, vzniká tlak na efektivní správu, ukládání a manipulaci s elektronickými dokumenty. Řada dokumentů v současné době začíná vznikat již v elektronické formě u původců (důležité evidence, zvukové a obrazové záznamy, fotografie apod.), také se provádí digitalizace stávajících fyzických dokumentů, aby se dochovalo svědectví o skutečnostech, které fyzické dokumenty obsahují, jelikož použitý materiál je citlivý a očekává se v dohledné době jeho dožití, nebo i za účelem zpřístupnění věrné podoby uložených archiválií uživatelské komunitě (badatelům) bez nutnosti používat původní dokument a tím snížit manipulaci s původním dokumentem.

Problematika střednědobého a dlouhodobého ukládání a archivace se zabývá daty a dokumenty, které byly v určitém okamžiku zafixovány jako neměnné. V případě dokumentů se jedná typicky o okamžik, kdy je připravovaný dokument prohlášen za hotový (např. po podpisu dokumentu není žádoucí provádět jakékoli jeho změny a měl by být v této podobě uchovávan).

V případě jiných dat je často tímto okamžikem myšlen okamžik vyexportování dat z jejich provozního systému. Systémově je třeba zajistit, aby ukládaný obsah zůstal neměnný.

Dlouhodobá archivace digitálních dokumentů vyžaduje trvalou péči, která vyžaduje určité náklady (větší než na pouhé uložení). Také příprava dokumentů pro uložení v dlouhodobém archivu představuje určité náklady. Minimálně je třeba k dokumentům připojit i příslušná metadata (popisná a technická) a dokumenty zabalit do balíčků vhodných pro archivaci.

Pro některá data mohou být náklady na jejich převedení do podoby vhodné pro dlouhodobou archivaci velmi vysoké, protože dosud neexistují dostupné metody vhodné pro tento převod. Příkladem může být dlouhodobá archivace složitějších databázových aplikací, která vyžaduje kromě archivace vlastních dat i dlouhodobou archivaci aplikačního softwaru. Na výzkumu a vývoji obecných metod pro archivaci složitějších komplexnějších dat se ve světě dosud pracuje.

Z těchto důvodů jsme problematiku ukládání a archivace rozdělili na dvě části:

1. Dlouhodobá archivace dokumentů,
2. Bezpečné dlouhodobé ukládání dat.

Dokumenty obvyklých (běžných) formátů (textové, obrazové, audio, video), pro které jsou nástroje na převedení do podoby vhodné pro dlouhodobou archivaci běžně k dispozici, navrhneme ukládat a dále spravovat v systému dlouhodobého archivu.

Data komplexnějšího charakteru, pro které by převod do archivních balíčků byl nákladný nebo v dané době nemožný, navrhneme pouze ukládat do digitálního úložiště. Uložení takovýchto dat do úložiště znamená odložení řešení problému dlouhodobé archivace na pozdější dobu. Přitom předpokládáme, že čas potřebný pro vyřešení metod pro dlouhodobou archivaci těchto dat bude kratší, než je životnost příslušného softwaru či příslušného formátu.

Rozdělením problematiky na dvě části máme možnost vyřešit trvalou archivaci jen těch dokumentů, u kterých je tento problém v současné době rozumně řešitelný a zároveň neztratit dokumenty a data u kterých trvalá archivace zatím není uspokojivě vyřešena.

Pro každý nový vstup (data či dokumenty), který bude nutno uložit, bude třeba nejprve specifikovat a vyhodnotit potřebnou délku uložení, formát dat a jeho potenciální ohrožení, existenci nástrojů pro převod dat do archivního formátu, hodnotu dat. Na základě těchto kritérií bude možno rozhodnout, zda má smysl data uložit do dlouhodobého archivu, nebo zda postačuje uložení v digitálním úložišti. (Uložení dat mimo dlouhodobý archiv přináší potenciální nebezpečí ztráty logické čitelnosti vlivem nedostupnosti software, kterým byla data pořízena, resp. kterým je možno jejich formát interpretovat).

Dlouhodobá archivace dokumentů

Při návrhu digitálního archivu je nutné mít na paměti neustálý vývoj technologií. Je třeba především volit řešení, které bude dostatečně otevřené, ověřené a podporované, a tím oddálí potřebu vyměnit technologie. Na druhou stranu, pokud bude nutné technologii vyměnit (dlouhodobě se s nutností výměny musí počítat, viz životní cyklus archivu), umožní použité řešení jednodušší přechod a migraci do nového prostředí počítačových technologií.

Pro budování digitálních archivů zaměřených na dlouhodobé uchování digitálních dokumentů se v současnosti ve světě vychází především ze standardu OAIS (Open Archival

Information System). Ten specifikuje základní funkční části otevřeného archivu, komunikaci s okolím, procesy a informační model ve formě informačních balíčků přijímaných, poskytovaných a především uložených v archivu. Podrobněji o modelu OAIS ve speciální kapitole. Z tohoto modelu vycházíme i při návrhu systému pro dlouhodobou archivaci digitálních dokumentů v rámci Technologického centra kraje. Podrobnější zdůvodnění použití modelu OAIS je uvedeno ve speciální kapitole.

Technologie pro provoz systému digitálního archivu, archivních úložišť a zpřístupňování informací se neustále mění a je nutné na tyto změny reagovat tak, aby systém digitálního archivu zůstal životaschopný. V rámci životního cyklu systému je nezbytné provádět pravidelný dohled nad morálním zastaráváním technického řešení. Je nezbytné mít připravené takové postupy, aby bylo možno dostatečně včas spustit proces obnovy včetně přípravy a finančního zajištění. V modelu OAIS se touto činností zabývá část nazývaná „Preservation planning“. Tato služba monitoruje změny vnějšího prostředí, které by mohly mít dopad na schopnost archivu chránit a udržet přístup k informacím v péči. Jako reakci na změny vytváří tato služba doporučení pro aktualizaci politik a procedur OAIS a pro přizpůsobení se těmto změnám (např. navrhuje změnit technický formát uložených souborů, který je ohrožen na jiný-nový formát. Připravuje a ověřuje i software pro migraci formátu.) Provozování této služby představuje jeden z podstatných nákladů na dlouhodobou archivaci.

Model OAIS je všeobecně odbornou veřejností přijímán i pro projekty Národní digitální archiv a Národní digitální knihovna, jejichž realizace se připravuje na celonárodní úrovni (nositeli projektů jsou Národní archiv ČR a Národní knihovna ČR). Oba tyto subjekty se budou dlouhodobě zabývat i činnostmi souvisejícími s „Preservation planning“. Podle názoru autorů této studie by se výstupy z jejich činnosti mohly využít i pro archivy na krajské úrovni. Obě instituce by tedy měly provádět sledování vývoje, přípravu SW, potřebného pro migraci doslouživších formátů, změny ukládacích technologií a vydávat pokyny a doporučení pro provedení migrací či změny technologie (např. typu média) pro dlouhodobou archivaci svých dokumentů. Tyto pokyny by měly být obdobné i pro dlouhodobou archivaci na krajské úrovni (KDS a KDR). Pokud by byl přijat tento postup „metodického řízení“ krajských archivů z NDA a NDK, mohly by se na krajské úrovni vydané pokyny a doporučení pouze zpracovat. To znamená například podle vydaného doporučení vybrat dokumenty s ohroženým formátem k migraci a tyto, přes migrační SW (získaný z NDA či NDK) převést do nového formátu a uložit je zpět do archivu. Tím by se náklady na digital preservation na krajské úrovni mohly podstatně zmenšit. Z tohoto důvodu doporučujeme najít či upravit pro vztah mezi Národním archivem, provozujícím Národní digitální archiv, Národní knihovnou provozující Národní digitální knihovnu a krajskými úřady provozujícími KDS a KDR rozumný právní statut, který by výše zmíněné metodické řízení umožňoval.

Problematika dlouhodobé archivace se dále věnuje dvěma kategoriím dokumentů:

1. Úřední dokumenty
2. Ostatní dokumenty kulturního dědictví.

Ad 1. Úřední dokumenty

Správa úředních dokumentů je stanovena příslušnými legislativními předpisy především zákonem č. 499/2004 Sb. a jeho novelizací zákonem č. 190/2009 Sb.

Novela Archivního zákona předepisuje původcům předávat digitální archiválie po uplynutí archivní lhůty do NDA. Národní standard pro elektronické spisové služby a vyhláška č. 191/2009 Sb. stanoví povinnost pro přenos dokumentů do budovaných digitálních archivů vytvořit přesně definované datové balíčky, které kromě vlastních dokumentů obsahují i velké množství metadat. Tyto balíčky budou akceptovány všemi typy archivů.

Nesporně nejjednodušším okamžikem pro vytvoření těchto balíčků je okamžik jejich výstupu z elektronické spisové služby (ERMS systému) a příjmu těchto dokumentů do KDS, protože v této době lze získat to nejširší množství metadat přímo ze spisové služby. Nelze racionálně předpokládat vůli, časové možnosti nebo dostatek podkladů pro vytváření těchto balíčků až po uplynutí skartačních lhůt, které se pohybují v rozmezí cca od 5 do 100 let.

Pokud vezmeme například dokument s archivační lhůtou A50, pak po 50ti letech od vzniku dokumentu se mají získat jeho metadata, sestavit balíček a předat do NDA. Otázky typu kde za 50 let bude možno metadata získat, v jakém formátu bude uchován originální dokument, zda bude ještě čitelný apod. možná nejsou nyní aktuální, ale systémový návrh řešení by měl nabídnout jejich řešení.

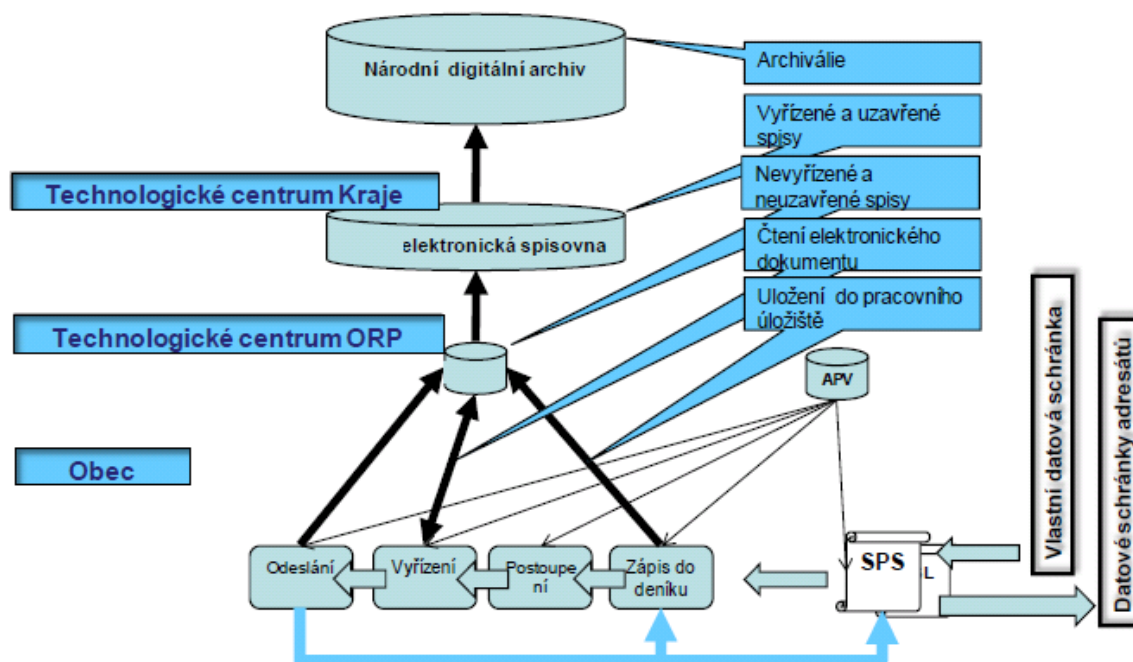
Použití standardu OAIS pro systém digitální spisovny se snaží uvedeným problémům předejít a sestavit informační balíčky s dokumenty ihned po uzavření dokumentu či spisu, kdy jsou veškeré potřebné údaje dostupné v aktuálních formátech a systémech. Takto připravený balíček se po 50 letech již jen do NDA přesune.

Úřední dokumenty a spisy vznikají a vyřizují se v různých IT systémech a aplikacích jako Podatelna, Elektronická spisová služba, Agendový systém pro stavební řízení, apod. Tyto systémy zajišťují příjem dokumentů, přípravu a vyřízení, odesílání a spojování do spisů v rámci správního řízení či jiných odborných procesů organizací. Závěrečná fáze těchto procesů se většinou nazývá uzavření dokumentů. Uzavřený dokument se již nesmí měnit a pro jeho uchování je třeba s ním zacházet předepsaným způsobem. Listinné dokumenty se předávají do listinných spisoven. Elektronické dokumenty a spisy se po uzavření ukládají do elektronické spisovny. Životnost dokumentů a spisů uložených v elektronické spisovně je řízena spisovým plánem organizace (původce). Uložené dokumenty a spisy čekají v elektronické spisovně na skartační řízení. Po uplynutí skartační (archivační) lhůty dojde buď ke skartaci dokumentů nebo dojde k výběru archiválií, které se předávají do nadřízeného digitálního archivu (např. Národní digitální archiv). Je třeba počítat s tím, že některé dokumenty mohou v elektronické spisovně zůstat po velmi dlouhou dobu, aniž by se skartovaly či předávaly. Analýzou spisového plánu Krajského úřadu zjistíme, že některé typy dokumentů mají skartační lhůty 50, 70 i více let.

Podobu uložení elektronického dokumentu musí systém elektronické spisovny zajistit ochranu uložených informací před ztrátou, důvěryhodnost uložených informací (nezměněnost a prokazatelnost vzniku v uvedeném čase) a čitelnost uložených informací v budoucnosti. Kromě toho musí elektronická spisovna zajistit i ochranu uložených informací proti neoprávněnému přístupu.

Na úrovni Kraje navrhujeme správu úředních dokumentů v době od uzavření (vyřízení) do vyřazení do Národního digitálního archivu zajistit Krajskou digitální spisovnou (KDS).

Krajská digitální spisovna navazuje na projekty implementace (či rozšíření) elektronické spisové služby u různých subjektů kraje. Propojení elektronických spisových služeb, Krajské digitální spisovny a následně i Národního digitálního archivu ukazuje následující obrázek.



Obrázek 3: Propojení ESS s KDS a NDA

Zdroj: Technologická centra krajů a obcí s rozšířenou působností, včetně spisových služeb (Koncept a východiska) - materiál MVČR http://www.egoncentrum.cz/index.php?option=com_rubberdoc&view=doc&id=46&format=raw&Itemid=29.

Ad 2. Ostatní dokumenty kulturního dědictví

Kromě dokumentů úředního významu, popsanych v předchozím odstavci, existuje na regionální úrovni celá řada dalších dokumentů, které nemají úřední charakter ani nevznikly činností nějakého úřadu. Jedná se například o cenné písemnosti a umělecká díla, fotografie, audio, video, časopisy a ostatní publikace vztahující se k regionu.

Podle způsobu vzniku se může jednat o výstupy z digitalizačních linek (to znamená, že po provedení digitalizace existuje ještě po nějakou dobu originální fyzický podklad) či o dokumenty vzniklé již primárně v digitální podobě. Vzhledem k tomu, že životnost původních fyzických originálů je často omezena (viz problematika tiskovin vytištěných na kyselém papíru) je třeba se k oběma těmto druhům chovat z pohledu dlouhodobé archivace stejně. Ztráta digitálního dokumentu z digitálního archivu je nenahraditelná.

Na úrovni Kraje navrhujeme ukládat tyto dokumenty v Krajském digitálním repozitáři (KDR).

Bezpečné dlouhodobé ukládání dat

Kromě dvou výše popsaných kategorií dat – dokumentů, pro které je vhodné použít technologii dlouhodobé archivace, existují na krajské úrovni ještě další data, jejichž uložení je třeba zajistit.

Jedná se např. o data ze zdravotní dokumentace krajských nemocnic, záznamy z pouličních kamer, logy z provozu síťových připojení (podle telekomunikačního zákona) a jakákoliv jiná data, která pocházejí z činnosti informačních systémů kraje a je třeba je z nejrůznějších důvodů střednědobě až dlouhodobě ochránit proti ztrátě.

Přitom použití technologie dlouhodobé archivace je pro tato data zbytečně nákladné či nedostupné, nebo potřeba jejich uložení není dlouhodobá. Dlouhodobost přitom není přesně definovaný termín, ale obecně ho lze chápat jako uložení nad 15-20 let.

Na úrovni Kraje předpokládáme ukládat tato data v Krajském digitálním úložišti (KDÚ).

V Krajském Digitálním Úložišti (KDÚ) budou uložena data nejrůznějšího typu především z toho důvodu, aby se včas zabránilo jejich zničení před definitivním rozhodnutím, jak s nimi bude naloženo později. Hlavním úkolem KDÚ je dlouhodobé bezpečné uložení počítačových dat spočívající v zabezpečení neměnnosti uložených dat ve formátu, v jakém byla uložena.

1.1.1 Návrh a popis architektury řešení

1.1.1.1 Požadavky na budoucí systém

Krajská digitalizační jednotka (KDJ)

Digitalizace se bude realizovat výhradně v Krajské Digitalizační Jednotce (KDJ).

KDJ se bude využívat převážně pro digitalizaci regionálního kulturního dědictví. Pro oblast úředních dokumentů Kraje se bude využívat pouze tehdy, nebude-li vytižena digitalizací regionálních kulturních dokumentů.

V KDJ se předpokládá zaměstnat jednoho až dva pracovníky v oblasti skenování a úprav obrazu, 1 pracovníka pro pořizování metadat, jednoho pracovníka pro přípravu dokumentů.

Základní pracoviště KDJ bude v budově Krajského úřadu, kde bude KDJ umístěna většinu na počátku projektu a pak i později dle potřeby. Součástí základního pracoviště bude i zamykací klimatizovaný mezisklad pro práci s historickými digitalizovanými fondy vybavený vysavačem pro odprášení a zamykacími vozíky na převoz dokumentů ke skeneru.

Požaduje se etapovité přemísťování KDJ do lokalit, kde se budou digitalizovat významné a vzácné listinné fondy, např. knihovna tepelského kláštera. S touto eventualitou se počítá i proto, aby při převozu listinných fondů nevyčísitelné ceny nemohlo dojít k jejich poškození.

Případné poškození KDJ při její přepravě (navíc může být přeprava pojištěna) je nesrovnatelné se škodou, která by mohla vzniknout při přepravě historicky cenných dokumentů na místo digitalizace.

Mobilita KDJ vyžaduje v místě digitalizace, např. v tepelském klášteře, vybudování této minimální infrastruktury:

- vhodné pracoviště pro umístění skenerů, pracovních stanic a serverů na zpracování digitálních obrazů a pracovního úložiště pro operativní uložení zabalených obrazů společně s jejich metadaty (t.zv. balíčky SIP) v areálu knihovny
 - připojení k internetu především pro vzdálenou komunikaci s "Registrem Digitalizace"

Následně uveden návrh portfolia možných typů skenerů, které jsou uvedeny v následujících tabulkách. Tabulky slouží jako analytický podklad pro odhad doby trvání digitalizace všech požadovaných fondů při různých kombinacích skenerů:

- robotický knižní skener (zvládnutí velkého množství běžných knih menších formátů, dvojstrana do A3)
- kamerový knižní skener (pro ruční skenování knih a větších formátů knih a dokumentů do A2)
- portálový velkoformátový kombinovaný skener formátu A1 (pro ruční skenování vzácných listin a s přídatnou kolébkou pro skenování velmi vzácných knih velkého rozměru)
- bubnový velkoformátový skener A0+ (pro plány a mapy)

Při výběru dokumentů pro digitalizaci se používají následující kritéria:

- historická dokumentární hodnota
- míra reálného nebo možného fyzického poškození
- frekvence využívání badatelskou veřejností

Krajem bylo rozhodnuto digitalizovat následující knižní fondy a periodika:

| Dokumenty k digitalizaci - Karlovarský kraj | | |
|---|-------------------|--------------------|
| aktualizace k 19.7.2010 | | |
| | počet knih | počet stran |
| muzeum Cheb | 150 | 15000 |
| muzeum Karlovy Vary | 3163 | 316300 |
| knihovna Jáchymov | 16 | 725 |
| knihovna Loket | 21 | 3624 |
| knihovna Cheb | 7 | 1622 |
| muzeum Sokolov | 23 | 5315 |
| Knihovna hum. věd Křest. ak. KV | 386 | 38600 |
| mezisoučet | 3766 | 381186 |
| KKKV balneo knihy | 1584 | 158400 |
| KKKV region knihy | 1500 | 15000 |
| celkem knihy | 5350 | 554586 |
| periodika | | |
| Balneologické listy | | 4896 |
| Balneologia Bohemica | | 3800 |
| Stráž míru | | 6900 |
| Arnika | | 2048 |
| celkem periodika | | 17644 |
| celkem knihy a periodika | | 572230 |
| Pozn.: u knih, kde nebylo uvedeno stránkování je počítáno průměrně 100 stran/1knihu | | |
| v současnosti nelze přesně zjistit počty: | | |
| | | cca |
| SVK Plzeň - dokumenty týkající se Karlovarského kraje | | 250000 |
| Klášteř Teplá | | 200000 |
| knihovna Střední keramické školy KV | | 50000 |
| místní periodika + regionální přílohy deníků | | 100000 |
| balneo (lékařská) periodika výběrově | | 100000 |
| muzejní sborníky | | 2000 |
| celkem stran - odhad | | 702000 |

Tabulka 3: Dokumenty k digitalizaci

Pro zadaný úkol byly navrženy digitalizační metriky jak podle různých druhů skenerů, tak podle dob zpracování a tyto metriky jak pro digitalizaci knih, tak pro digitalizaci periodik byly uloženy do následujících dvou tabulek.

Upozorňujeme, že následující metriky jsou pouze řádové odhady (nejsou uváděny skenery konkrétních výrobců ani konkrétní zpracovatelský SW pro image), ale vychází se z praktických zkušeností s různými typy skenerů a s následujícími potřebnými úkony, z nichž některé je možno zautomatizovat díky SW na zpracování image, některé závisí pouze na rychlosti lidské práce a její preciznosti.

Úkony, které souvisí s digitalizací knih či periodik jsou následující:

- příprava dokumentů k digitalizaci (není zahrnuto do časů v tabulce):
 - odprášení
 - doprava ke skenerům a od skenerů
- vlastní skenování a uložení image do operativního úložiště
 - čas nezávisí přímo na technické době skenu dvojstrany, ale též na ručních operacích, které se skenováním souvisí
- doplňování metadat bezprostředně před nebo po skenování knihy či jednoho vydání periodika
- úprava naskenovaných image (dle použitého SW od zcela ručního postupu až k postupu zcela automatickému). Jedná se o následující úkony, které se mohou provádět ručně či automaticky, pro každou dvojstranu samostatně, či s jedním nastavením pro celou knihu či periodikum:
 - splitting = půlení naskenované dvojstránky
 - skewing = úhlové natočení šikmo vložené stránky
 - cropping = ořez stránky
 - correcting levels = nastavení barevných a jasových úrovní
- zabalení image i metadat do jednoho balíčku SIP (Submission Information Package) a zaslání do KDR (může se dle použitého SW provádět ručně, poloautomaticky, automaticky, jednotlivě či dávkově)

| digitalizační metriky knih | skener 1 | skener 2 | skener 3 | skener 4 | požadovaný počet stran | |
|--|-------------------------|------------------------|-------------------------|----------------|------------------------|------------------------|
| | | | | | hrubý odhad celkem | přesně zjištěno celkem |
| <i>průměr 100 str. / knihu</i> | | | | | | |
| typ skeneru | robotický knižní skener | kamerový knižní skener | portálový knižní skener | bubnový skener | | |
| maximální formát dvojstrany | A3 | A2 | A1 | A0+ | | |
| průměrná doba skenování stránky [s] | 4,5 | 30 | 90 | | | |
| průměrná doba pořízení metadat ke knize v přepočtu na stránku [s] | 3 | 3 | 3 | | | |
| úprava naskenovaných image v přepočtu na stránku [s] | 3 | 3 | 3 | | | |
| zabalení do ukládacího formátu a předání k uložení v přepočtu na stránku [s] | 3 | 3 | 3 | | | |
| průměrná doba zpracování 1 stránky [s] | 14 | 39 | 99 | | | |
| počet digitalizovaných stránek / hodinu | 267 | 92 | 36 | | | |
| počet digitalizovaných stránek / směnu | 1 600 | 554 | 218 | | | |
| počet digitalizovaných stránek / rok (jednosměnný provoz) | 400 000 | 138 462 | 54 545 | | 806 586 | 554 586 |
| počet digitalizovaných dokumentů / rok (dvojsměnný provoz) | 800 000 | 276 923 | 109 091 | | | |
| hodin / směnu | 6 | 6 | 6 | | | |
| směn / rok (jednosměnný provoz) | 250 | 250 | 250 | | | |
| směn / rok (dvojsměnný provoz) | 500 | 500 | 500 | | | |

Tabulka 4: Digitalizační metriky knih

| digitalizační metriky periodik | skener 1 | skener 2 | skener 3 | skener 4 | požadovaný počet stran | |
|--|-------------------------|------------------------|-------------------------|----------------|------------------------|-----------------|
| | | | | | hrubý odhad | přesně zjištěno |
| <i>průměr 10 str. / periodikum</i> | | | | | | |
| typ skeneru | robotický knižní skener | kamerový knižní skener | portálový knižní skener | bubnový skener | | |
| maximální formát dvojstrany | A3 | A2 | A1 | A0+ | | |
| průměrná doba skenování stránky [s] | | 60 | 90 | 300 | | |
| průměrná doba pořízení metadat ke stránce [s] | | 30 | 30 | 30 | | |
| úprava naskenovaných image v přepočtu na stránku [s] | | 30 | 30 | 30 | | |
| zabalení do ukládacího formátu a předání k uložení v přepočtu na stránku [s] | | 30 | 30 | 30 | | |
| průměrná doba zpracování 1 stránky [s] | | 150 | 180 | 390 | | |
| počet digitalizovaných stránek / hodinu | | 24 | 20 | 9 | | |
| počet digitalizovaných stránek / směnu | | 144 | 120 | 55 | | |
| počet digitalizovaných dokumentů / rok (jednosměnný provoz) | | 36 000 | 30 000 | 13 846 | 467 644 | 17 644 |
| počet digitalizovaných stránek / rok (dvojsměnný provoz) | | 72 000 | 60 000 | 27 692 | | |
| hodin / směnu | | 6 | 6 | 6 | | |
| směn / rok (jednosměnný provoz) | | 250 | 250 | 250 | | |
| směn / rok (dvojsměnný provoz) | | 500 | 500 | 500 | | |

Tabulka 5: Digitalizační metriky periodik

Vzhledem požadované skladbě a počtu knih a periodik navrhujeme tuto základní variantu řešení, která vychází z kombinace digitalizace vlastními silami v rámci KDJ a digitalizace za použití outsourcingu externí firmou. Vlastními silami budou

digitalizovány menší formáty – do formátu A2. Digitalizace větších formátů, kterých je relativně velmi málo, bude realizováno externí firmou.

Digitalizovat se bude v jedné směně na jednom skeneru:

- 1 kamerový knižní skener (pro ruční skenování menších i větších formátů knih a dokumentů do formátu dvojstrany A2)

Uvažujeme-li výše uvedené metriky a budeme digitalizovat vlastními silami, v hrubém odhadu docházíme k následujícím závěrům:

- **do 6ti let se může zdigitalizovat celý požadovaný knižní fond, do dalších 13ti let se na stejném skeneru mohou zdigitalizovat veškerá požadovaná periodika. Hrubý odhad celkové doby digitalizace všech fondů může činit 19 let**
- **digitalizační jednotka se dle priorit může přesouvat mezi lokalitami a tam skenovat střídavě knihy či periodika, opět dle priorit**
- **je samozřejmé, že konkrétní skener bude nutno po době jeho životnosti nahradit jiným, podobného typu, ale to již bude po době udržitelnosti projektu**
- **zjistí-li se, že v nějaké lokalitě je též větší počet velkoformátových dokumentů, požádá se externí firma o digitalizaci těchto dokumentů přímo v této lokalitě**
- **zjistí-li se, že v nějaké lokalitě je pouze malý počet velkoformátových dokumentů, dopraví se tyto dokumenty přímo na digitalizační pracoviště externí firmy a zpět**

Krajská digitální spisovna (KDS)

Navrhujeme vybudovat KDS podle modelu OAIS.

Základní požadavky na vybudování KDS jsou následující:

- Vybudování KDS jako součásti TC Kraje v souladu s požadavky kladenými na důvěryhodnou instituci a stanovení a plnění podmínek udržitelnosti provozu v současnosti i budoucnosti
- Vlastníkem dokumentů uložených v KDS je příslušný původce. Kraj je provozovatelem KDS, s původci uzavírá smlouvy a původce metodicky řídí
- Systémové řešení dlouhodobého uchovávání a zpřístupňování úředních elektronických dokumentů vytvořených určenými původci
- Nepřetržitě zajištění základních charakteristik elektronických dokumentů uložených v KDS – životaschopnosti, čitelnosti, pochopitelnosti, autentičnosti a identifikace
- Z důvodu zajištění naprosté nezpochybnitelnosti právní průkaznosti uložených dokumentů zesiluje se prokázání jejich autentičnosti opakovaným pravidelným přerazítkováním otisků balíčků AIP časovým razítkem autorizované časové autority ještě v době platnosti původního elektronického podpisu dokumentů či posledního minulého časového razítka autorizované časové autority
- Přebírání, zpracování, archivace a zpřístupňování úředních dokumentů ze systémů elektronických spisových služeb a ERMS systémů smluvních původců na území Kraje

- Možnost provádění skartačního řízení v KDS a to tak, že každý původce provádí vzdáleně skartační řízení nad svými dokumenty, které má v KDS uloženy
- Možnost vzdáleného přístupu archiváře Národního digitálního archivu (NDA) při skartačních řízeních jednotlivých původců s možností přímého exportu vybraných AIP do NDA

Původci KDS jsou současně i konzumenty služeb poskytovaných KDS

1.1.1.1.1.1 Identifikace a kategorizace dokumentů, metadata

Podle Národního standardu pro vedení elektronické spisové služby je každý původce, který je povinen vést elektronickou spisovou službu, povinen opatřit každý dokument unikátním identifikátorem. Hodnota tohoto identifikátoru bude předávána v metadatech dokumentu.

Pokud bude takový původce předávat spisy či dokumenty do Krajské digitální spisovny, bude muset jejich typy povinně zařadit do svého jednotného spisového plánu. (Je povinností původce přidat do používaného spisového plánu další položky tak, aby veškeré produkované či evidované dokumenty bylo možno do jednotného spisového plánu zařadit.)

Každý původce si může svůj spisový plán sestavit podle svých potřeb při respektování příslušných doporučení. Spisový plán je vhodné konzultovat s příslušným archivem. Krajská digitální spisovna bude od původců, kteří jí předávají dokumenty a spisy k uložení, jejich spisové plány (příp. obdobné vnitřní předpisy) vyžadovat.

Krajská digitální spisovna bude evidovat veškeré spisové plány všech původců, se kterými bude mít smlouvu o uložení jejich dokumentů. Tito původci budou mít povinnost aktualizovat své spisové plány v krajské spisovně tak, aby veškeré dokumenty byly do příslušného spisového plánu zařaditelné. Zařazení do spisového plánu bude na vstupu dokumentů kontrolováno vstupní procedurou. Dokumenty, které nebudou kategorizovány podle aktuálního spisového plánu, budou vstupní procedurou odmítnuty. Původce musí být na tuto skutečnost upozorněn.

Povinná katalogizace podle spisového plánu vyplývá z Národního standardu pro vedení elektronické spisové služby. Navíc zajistí možnost vyhledávání dokumentů různých původců podle této indexace a také řádné provádění skartačních řízení.

Dokumenty budou do Krajské digitální spisovny předávány ve formátu definovaném Národním standardem pro vedení elektronické spisové služby pro předávání dokumentů do archivu (viz příloha 1, část 2 tohoto standardu). Výstupním datovým formátem metadat, jimiž jsou podle vyhlášky opatřovány dokumenty, se rozumí formát XML podle schématu XML. Celý standard je k dispozici na <http://www.mvcr.cz/clanek/narodni-standard-pro-elektronicke-systemy-spisove-sluzby.aspx>.

1.1.1.1.1.2 Formáty elektronických dokumentů

Formáty elektronických dokumentů uložených v KDS musí odpovídat aktuální vyhlášce MV ČR, která určí výstupní formáty systémů spisových služeb. Současná vyhláška č. 191/2009 Sb., o podrobnostech výkonu spisové služby stanovuje tyto povolené typy:

- Jako výstupní datový formát statických textových dokumentů a statických kombinovaných textových a obrazových dokumentů se použije:

- Formát Portable Document Format/Archive (PDF/A, ISO19005) se použije jako výstupní datový formát statických textových dokumentů a statických kombinovaných textových a obrazových dokumentů.
- Jako výstupní datové formáty statických obrazových dokumentů se použijí:
 - formát Portable Network Graphics (PNG, ISO/IEC 15948)
 - formát Tagged Image File Format (TIFF, revize 6 - nekomprimovaný)
 - formát JPEG File Interchange Format (JPEG/JFIF, ISO/IEC 10918)
- Jako výstupní datové formáty dynamických obrazových dokumentů se použijí:
 - video programový nástroj pro komprimaci dat (kodek) Moving Picture Experts Group Phase 2 (MPEG-2, ISO/IEC 13818)
 - video programový nástroj pro komprimaci dat (kodek) Moving Picture Experts Group Phase 1 (MPEG-1, ISO/IEC 11172)
 - formát Graphics Interchange Format (GIF)
- Jako výstupní datové formáty zvukových dokumentů se použijí:
 - zvukový programový nástroj pro komprimaci dat (kodek) MP2 (MPEG-1 Audio Layer 2)
 - zvukový programový nástroj pro komprimaci dat (kodek) MP3 (MPEG-1 Audio Layer 3)
 - formát Waveform audio format (WAV), modulace Pulse-code modulation (PCM)

1.1.1.1.1.3 Dimenzování KDS

Původci KDS jsou subjekty, které používají či budou používat systémy elektronické spisové služby ESS.

Předpokládáme, že v budoucnosti budou mít i velmi malé obce či zřizované organizace k dispozici jednoduché hostované systémy ESS. Důvodem je celostátně zavedený systém datových schránek, kterými prochází či bude procházet velká část korespondence či dokumentů v elektronické podobě.

Zdrojem "digital born" elektronických dokumentů jsou jednak dokumenty doručené či odeslané systémem datových schránek, jednak interní dokumenty přímo elektronicky vytvořené v agendových systémech původců a následně zaevidované a uložené v ESS.

Původci elektronických dokumentů tedy jsou nebo potenciálně budou následující subjekty:

- orgány Kraje (hejtman, Rada, Zastupitelstvo, Krajský úřad včetně jeho organizačních jednotek)
- obce v rámci Kraje
- zřízené organizace Krajem či obcemi Kraje

Následující tabulka odhaduje současné průměrné roční datové přírůstky do KDS z elektronických spisových služeb na území Kraje. Vychází přitom ze zjištěných údajů, kvalifikovaných odhadů a metrik.

Do odhadu nejsou vůbec zahrnuty potenciální velké možné budoucí přírůstky "digital born" dokumentů typu elektronických fotografií a především elektronických audiovizuálních dokumentů, např. zasedání zastupitelstev, různých interview, dokumentárních záběrů z nejrůznějších oblastí veřejné správy, které jistě budou přílohami řady dokumentů. Tyto budoucí potřeby mohou **řádově** zvýšit nároky na objem úložiště.

| Odhady ročních přírůstků do KDS dle původců | | | | |
|--|--|---|------------------------------|-------------------------------------|
| <i>metrika: dokument = průměrně 2,3 stránky formátu A4</i> | | | | |
| <i>metrika: elektronická stránka A4 = průměrně 80 kB</i> | | | | |
| <i>metrika: kvalitně naskenovaná barevná příloha A4 = průměrně 12 MB</i> | | | | |
| původci | průměrný počet dokumentů /původce | průměrný počet stránek / původce | počet původců v Kraji | roční přírůstky ve stránkách |
| Krajský úřad | 300 000 | 690 000 | 1 | 690 000 |
| obec typu PO1 | 710 | 1 633 | 117 | 191 061 |
| obec typu PO2 | 6 970 | 16 031 | 7 | 112 217 |
| obec typu PO3 | 63 800 | 146 740 | 7 | 1 027 180 |
| mateřská škola | | 300 | 36 | 10 800 |
| zákl. škola | | 4 000 | 75 | 300 000 |
| ostatní organizace | | 100 | 13 | 1 300 |
| celkový roční přírůstek stránek A4 v elektronické podobě | | | | 2 332 558 |
| z toho 90% běžné dokumenty | | | | 2 099 302 |
| z toho 10% kvalitní přílohy či fotografie | | | | 233 256 |
| celkový roční datový přírůstek [TB] | | | | 2,967014 |

Tabulka 6: Roční přírůstek do KDS

Vzhledem ke skartačním lhůtám a opakovaně probíhajícímu skartačnímu řízení neroste objem KDS do nekonečna, ale existuje určitá závislost mezi velikostí ročního přírůstku a potřebné maximální kapacity digitální spisovny.

| Přírůstky a úbytky ze spisovny v [%] z ročního pravidelného přírůstku 100% | | | | | | | | | | | | |
|--|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| rok | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| přírůstky | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| úbytky | | | | -5 | | -40 | | | | | -50 | |
| úbytky | | | | | -5 | | -40 | | | | | -50 |
| úbytky | | | | | | -5 | | -40 | | | | |
| úbytky | | | | | | | -5 | | -40 | | | |
| úbytky | | | | | | | | -5 | | -40 | | |
| úbytky | | | | | | | | | -5 | | -40 | |
| úbytky | | | | | | | | | | -5 | | -40 |
| úbytky | | | | | | | | | | | -5 | |
| úbytky | | | | | | | | | | | | -5 |
| potřebná kapacita spisovny v [%] z ročního přírůstku 100% | 100 | 200 | 300 | 395 | 490 | 545 | 600 | 655 | 710 | 765 | 770 | 775 |

Tabulka 7: přírůstky a úbytky KDS

Při konstrukci výpočtu potřebné kapacity elektronické spisovny vycházíme ze statistiky Národního archivu, která uvádí procentuální poměr dokumentů, které jsou v jednotlivých létech ve spisovných skartovány či předávány do státních archivů.

- po 3 letech se skartuje či předává 5% dokumentů
- po 5 letech se skartuje či předává 40% dokumentů
- po 10 letech se skartuje či předává 50% dokumentů
- 5% dokumentů má delší skartovací lhůtu

Zadáme-li tyto údaje do tabulky, zjistíme srovnání přírůstkové křivky na konečný lineární přírůstek 5%/rok. Tento jev nastává v 10. roce na kapacitě spisovny a činí 765% pravidelného ročního přírůstku.

S přihlédnutím k tabulce doporučujeme spisovnu dimenzovat na 5 let s kapacitou 500% pravidelného ročního přírůstku.

Ve čtvrtém roce doporučujeme rozšířit archivní úložiště spisovny na další potřebnou kapacitu podle aktuální situace a odhadů do budoucnosti, z dnešního hlediska o dalších 400% pravidelného ročního přírůstku.

Rozšiřování kapacity úložiště je výhodné realizovat v dobře uvážených krocích, protože cena úložiště vyjádřená v Kč/GByte neustále klesá.

Vzhledem k výše uvedenému odhadu ročních přírůstků předpokládáme:

Zřídit rozšiřitelné úložiště KDS, přičemž

- **výchozí kapacitu dimenzujeme na 5 let, což je cca 3 TB * 5 = 15 TB**
- **ve třetím roce rozhodneme, jak velkou kapacitu budeme dokupovat, ze současného pohledu např. dalších 12 TB**

Krajský digitální repozitář (KDR)

Požadujeme vybudovat KDR podle modelu OAIS.

Základní požadavky na vybudování KDR jsou následující:

- Vybudování KDR jako součásti TC Kraje v souladu s požadavky kladenými na důvěryhodnou instituci a stanovení a plnění podmínek udržitelnosti provozu v současnosti i budoucnosti
- Vlastníkem všech dokumentů uložených v KDR je Kraj
- Systémové řešení trvalého uchování a zpřístupňování elektronických dokumentů kulturního dědictví z regionu Kraje. Jedná se i o digitalizované dokumenty regionálního kulturního dědictví
- Nepřetržitě zajištění základních charakteristik elektronických dokumentů uložených v KDS – životaschopnosti, čitelnosti, pochopitelnosti, autentičnosti a identifikace
- Z důvodu uložení archivních dokumentů regionálního kulturního dědictví nemá zajištění naprosté nezpochybnitelnosti právní průkaznosti uložených dokumentů smysl a autentičnost uložených digitálních objektů a případných operací s nimi se prokazuje dostatečným způsobem zápisy v uchovávacích metadatech v příslušných balíčcích AIP a v Systému správy dat (Data Management)
- Výběr a trvalé uchování původně elektronických či digitalizovaných dokumentů kulturního dědictví regionálního významu od libovolných původců v regionu Kraje po dohodě s nimi a vytváření příslušných fondů či sbírek
- Možnost provádění vnitřní skartace v KDR, což znamená výmaz těch balíčků AIP, které se do repozitáře dostaly omylem. Při vnitřní skartaci se smažou vlastní digitální dokumenty, nikoliv jejich metadata
- Možnost meziarchivní výměny dokumentů podle dohody s jinými digitálními repozitáři či archivy s preferencí přímého importu / exportu archivních informačních balíčků AIP

Původci KDR jsou:

- subjekty na území Kraje, u nichž vznikají digitální dokumenty regionálního kulturního dědictví

Konzumenty služeb KDR jsou subjekty mající zájem o regionální kulturní dědictví Kraje:

- badatelé

- široká kulturní veřejnost

1.1.1.1.1.4 Identifikace a kategorizace dokumentů, metadata

Jednoznačnou identifikaci dokumentů – digitálních objektů – zajišťuje po jejich vstupním zpracování Systém správy dat KDR, který tento jednoznačný identifikátor vygeneruje a uloží jak do své databáze, tak současně i s ostatními metadaty a samotným dokumentem do archivního informačního balíčku AIP.

Pro uživatelské vyhledávání dokumentů se využijí popisná metadata, která mohou být koncipována podle následujících standardů či celostátních nebo mezinárodních doporučení:

- standard metadat stanovený Národní knihovnou pro knihovní systémy
- možné vazby na číselníky stanovené Národní knihovnou (např. Celostátní databáze národních autorit vedená Národní knihovnou)
- základní archivní metadata používaná při budování archivních fondů a sbírek
- potencionálně může vzniknout požadavek i na využití standardů používaných v muzejnictví

1.1.1.1.1.5 Formáty elektronických dokumentů

Povolené formáty elektronických dokumentů pro uložení v KDR budou převážně stejné jako formáty pro KDS. Seznam povolených formátů lze pro KDR v podstatě libovolně rozšiřovat. Je třeba si ale uvědomit, že s každým povoleným formátem se zvětšují náklady na dlouhodobé uložení.

1.1.1.1.1.6 Dimenzování KDR

Při dimenzování archivního úložiště pro KDR vycházíme z požadovaného množství digitalizovaných knih a periodik regionálního kulturního dědictví včetně rozumné rezervy s výhledem na případnou další digitalizaci či ukládání již tzv. "digital born" dokumentů, které budou případně v této oblasti přímo digitálně vznikat.

Vzhledem ke stále se snižujícím cenám paměťových médií (poměr cena / jednotku datového objemu) požadujeme archivní úložiště s možností postupného etapovitého rozšiřování úložné kapacity.

Dle našeho odhadu bude v průběhu 19 let digitalizováno cca 807 tisíc knižních stránek a 468 tisíc stránek periodik, což dohromady činí 1 275 000 stránek.

Předpokládáme, že průměrná digitalizovaná stránka bude průměrem mezi formátem A4 a A3, čili průměrným mixem A4 a A3.

Pro výpočet požadovaného datového objemu je třeba stanovit celkovou naskenovanou plochu. Vyjděme z následující tabulky:

| Metriky plošných formátů | | |
|--------------------------|-------------------|------------------------------------|
| formát | počet A4 / formát | plocha formátu [m ²] |
| A4 | 1 | A0/16 |
| A3 | 2 | A0/8 |
| A2 | 4 | A0/4 |
| A1 | 8 | A0/2 |
| A0 | 16 | 1 |
| mix (A4+A3) | 1,5 | průměr (1/16; 1/8) = 0,09375 |

Tabulka 1: Metriky plošných formátů

Dle této metriky má celková skenovaná plocha velikost 0,09375 * 1 275 000

= cca 120 000 m²

Při výpočtu datového objemu úložiště musíme vycházet i z kvality skenované plochy.

Předpokládáme-li, že nám u většiny dokumentů postačí rozlišení 300 dpi a zhruba u poloviny dokumenty je nutno skenovat v barvě a u poloviny nám postačí odstíny šedi, dojdeme s použitím následující tabulky k níže uvedeným výsledkům pro objemy úložišť:

| Veškeré možné varianty kvality digitalizace dokumentů a z toho vyplývajícího datového objemu | | | | | | | | | |
|--|----------------------|----------|-------------------------------------|----------|--------------------------------|----------|---------------------|--------------|--------|
| Výpočet objemu | | | | | | | potřebný objem [TB] | | |
| Objemové metriky | nekomprimováno 100% | | JPEG 2000 bezestr. MC kompr. na 48% | | JPEG ztrátová UC kompr. na 10% | | Master Copy MC | User Copy UC | celkem |
| barevná hloubka čb - 1 bit | [MB/m ²] | [MB/A4] | [MB/m ²] | [MB/A4] | [MB/m ²] | [MB/A4] | | | |
| 300 dpi | 17,43753 | 1,087579 | 8,370017 | 0,522038 | 1,743753 | 0,108758 | 1,00 | 0,21 | 1,21 |
| 400 dpi | 31,00006 | 1,933474 | 14,88003 | 0,928067 | 3,100006 | 0,193347 | 1,79 | 0,37 | 2,16 |
| 600 dpi | 69,75014 | 4,350316 | 33,48007 | 2,088152 | 6,975014 | 0,435032 | 4,02 | 0,84 | 4,85 |
| barevná hloubka šedivá - 8 bit | [MB/m ²] | [MB/A4] | [MB/m ²] | [MB/A4] | [MB/m ²] | [MB/A4] | | | |
| 300 dpi | 139,5003 | 8,700632 | 66,96013 | 4,176304 | 13,95003 | 0,870063 | 8,04 | 1,67 | 9,71 |
| 400 dpi | 248,0005 | 15,46779 | 119,0402 | 7,42454 | 24,80005 | 1,546779 | 14,28 | 2,98 | 17,26 |
| 600 dpi | 558,0011 | 34,80253 | 267,8405 | 16,70521 | 55,80011 | 3,480253 | 32,14 | 6,70 | 38,84 |
| barevná hloubka barva - 24 bit | [MB/m ²] | [MB/A4] | [MB/m ²] | [MB/A4] | [MB/m ²] | [MB/A4] | | | |
| 300 dpi | 418,5008 | 26,1019 | 200,8804 | 12,52891 | 41,85008 | 2,61019 | 24,11 | 5,02 | 29,13 |
| 400 dpi | 744,0015 | 46,40337 | 357,1207 | 22,27362 | 74,40015 | 4,640337 | 42,85 | 8,93 | 51,78 |
| 600 dpi | 1674,003 | 104,4076 | 803,5216 | 50,11564 | 167,4003 | 10,44076 | 96,42 | 20,09 | 116,51 |
| zadaná hodnota [m ²] | 120 000 | | | | | | | | |

Tabulka 8: Varianty kvality digitalizace

Do samotného archivního úložiště KDR požadujeme ukládat tzv. "Master Copy" skenů v bezztrátovém formátu JPEG 2000.

Celková požadovaná kapacita KDR je tedy průměrem 8,04 TB (vše v šedi) a 24,11 (vše v barvě), což činí cca 16 TB. Vzhledem k potřebné rezervě a předem těžko odhadnutelným přírůstkům "digital born" dokumentů doporučujeme celkovou kapacitu do 19 let dimenzovat na 30 TB. To ovšem neznamená, pořizovat tuto kapacitu okamžitě.

Pro KDR požadujeme dokupovat paměťovou kapacitu po etapách, se startovací kapacitou 4 TB, s výhledem do 19 let na rozšíření do 30 TB.

Jako systém zpřístupnění využíváme v této oblasti systém KRAMERIUS. Do jeho úložiště – nejedná se o archivní systém! - pouhý filesystém, který má svou zálohu již v KDR - požadujeme ukládat t.zv. "User Copy" skenů ve ztrátovém formátu JPEG. Paměťové nároky na toto úložiště jsou celkem 3,4 TB. Zde doporučujeme startovací kapacitu 0,5 TB s dalším rozšířením dle potřeb.

Krajské digitální úložiště (KDÚ)

Zdrojem dat pro Krajské digitální úložiště mohou být různé informační systémy provozované u různých původců.

Na rozdíl od problematiky archivace dokumentů uvedené v předchozích dvou kapitolách se KDU zabývá uložením předem obecně nespécifikovaných dat, kdy specifikace dat, jejich metadat a šablony pro uchovávání a proces akvizice dat budou tvořeny vždy konkrétně dle požadavku na uložení dat.

Studie předpokládá uložení archivních dat v KDÚ ve formě souborů ukládaných původci prostřednictvím určených standardních souborových systémů, prostřednictvím API ukládajícího HW/SW původce nebo prostřednictvím aplikací umístěných v TC kraje (využití úložiště vrstvy Tier 3 dle „Studie proveditelnosti hostingového centra Kraje“). Úložiště neslouží pro přímé ukládání provozních dat původců na blokové úrovni, ale může sloužit pro přímé ukládání zálohovacích dat původců.

Příkladem jsou soubory - logy z provozních systémů. Ty se pomocí některého standardního souborového protokolu (ftp, nfs apod.) dostupného v operačním systému přenesou po síti a uloží se do adresářové struktury na disk.

Dalším příkladem je ukládání zdravotnické dokumentace kompatibilní s PACS (Picture Archiving and Communication System) systémy. Tento typ zdravotnické dokumentace je výstupem z jednotlivých digitálních lékařských přístrojů (modalit) schopných předávat obrazová data v různých formátech, z nichž nejvýznamnější jsou DICOM, HL7 (případně i národní standard DASTA).

Na různých úrovních se budují tzv. PACS archivy, což jsou specializované systémy pro dlouhodobou archivaci a vyhledávání obrazových dat. Takový archiv je určen pro křížové zálohování obrazových dat nebo k rozšíření vlastních archivačních kapacit nemocnice. Centralizované řešení na úrovni kraje může redukovat investiční a hlavně provozní náklady v jednotlivých zdravotnických zařízeních.

1.1.1.1.1.7 Odhad množství dat

Metriky použité pro odhad:

Zdravotnická dokumentace:

Velikost dat zdravotnické dokumentace je značně variabilní. To dokumentuje následující tabulka s přehledem objemu datových výstupů z různých zařízení.

| Zařízení (Modality) | Typ dat | Odhad velikosti dat na jedno vyšetření |
|------------------------------|---------------------------|--|
| Cardiac angiography | Motion video (monochrome) | 400 MB |
| Cardiac angiography with PCI | Motion video (monochrome) | 1,000 MB (1 GB) |
| Hemodynamics | Waveforms | 0.03 MB |
| Electrophysiology study | Waveforms | 250 MB |
| Intravenous ultrasound | Motion video (monochrome) | 10–750 MB |
| Echocardiology | Motion video (color) | 60 MB |
| EKG | Waveforms | 0.05 MB |
| PET | Still-frame image series | 30 MB |
| 2D Mammogram | Still-frame image series | 120 MB |
| CT | Still-frame image series | 160 MB |
| 3D MRI | Still-frame image series | 150 MB |
| 2D X-Ray | Still-frame image series | 30 MB |
| 3D CT 500 Slice series | Still-frame image series | +1,000 MB (+1 GB) |

Tabulka 9: Objemy zdravotní dokumentace

Source: RCG Healthcare Consulting, Massachusetts General Hospital

Proto odhad velikosti potřebného úložiště nelze na této úrovni provést. Jediným vodítkem pro odhad velikosti zdravotnické dokumentace jsou zkušenosti z provozu. Vycházíme z objemu realizovaném v podobně velkém kraji, požadovaná velikost je 8 TB na 5 let.

Provozní logy:

Odhadujeme objem 0,5 TB za rok, celkem za 5 let 2,5 TB.

Audio, video

Požadavky ORP (z dotazníku): 1 TB x 9 (7 ORP a 2 další velké obce) jako základ. Celková kapacita tedy cca. 15 TB za 5 let.

| Oblast | Objem dat za 5 let |
|--------------------------|--------------------|
| Zdravotnická dokumentace | 8 TB |
| Logy | 2,5 TB |
| Audio, video | 15 TB |

Tabulka 10: Objem dat pro KDÚ za 5 let ročně

1.1.1.1.1.8 Identifikace a kategorizace dat, metadata

V případě KDU nelze předem specifikovat žádná metadata. Přesto řada formátů (například formáty zdravotnické dokumentace DICOM a HL7) metadata obsahují.

Jediným požadavkem na metadata KDU jsou metadata pro vytvoření popisu každého logického segmentu KDU v katalogu.

1.1.1.1.1.9 Formáty dat

Pro KDU nelze předem stanovit žádná pravidla pro formáty dat.

1.1.1.1.1.10 Souhrnný odhad velikosti fyzického úložiště pro KDÚ

Pro úložiště KDU nelze v otázce skartace dat zatím předikovat. Hypotetickým předpokladem může být, že se po pěti letech skartuje 30 procent dat.

Dále předpokládáme, že se hodnoty po celou dobu nemění a že na počátku nebude třeba uložit velký objem stávajících dat.

Je třeba počítat s neustálým nárůstem objemu dat úložiště a proto je třeba volit pro jeho realizaci zařízení s velkou škálovatelností a možností postupného rozšiřování¹.

Pravidelné sledování přírůstku objemu dat a plánování velikosti úložiště patří k základním úlohám správce úložiště.

Podmínkou je, aby zvolený souborový systém byl určený pro archivaci obsahu a byl dimenzován pro uložení potřebného objemu dat (desítky až stovky TB).

KDÚ je úložištěm určeným pro data aplikací, ze kterých není možno vytvářet vstupní informační balíčky (SIP), protože metadata nejsou známa či jsou integrována do vlastních dat. Dalším důvodem může být fakt, že použití technologie dlouhodobé archivace by bylo pro tato data zbytečně nákladné či nedostupné, nebo potřeba jejich uložení není dlouhodobá.

Základní požadavky na vybudování KDÚ jsou následující:

- Vybudování KDÚ jako součásti TC Kraje v souladu s požadavky kladenými na důvěryhodnou instituci a stanovení a plnění podmínek udržitelnosti provozu v současnosti i budoucnosti
- Vlastníkem dat uložených v KDÚ je příslušný původce. Kraj je provozovatelem KDÚ, s původci uzavírá smlouvy o uložení. Pokud je kraj zřizovatelem původce je i vlastníkem dat.
- Systémové řešení služby krátkodobého až střednědobého uchování dat aplikací pro smluvní původce působící na území Kraje
- Evidence jednotlivých typů dat (logických segmentů) v katalogu
- Vstup dat a přístup k datům bude vždy přes uživatelský interface. Aplikační rozhraní do KDU se nebude vyvíjet.

¹ Postupné rozšiřování je výhodné nejen z pohledu ceny (tak jak cena na uložení jednoho MB dat postupně klesá), ale také proto, že technologie úložišť se neustále vyvíjí. Je třeba počítat s tím, že jednou bude třeba vyměnit celou ukládací platformu, protože skončí její morální životnost.

Požaduje se umožnit přístup k datům KDÚ i jiným subjektům nežli jsou původci a to na základě konkrétně definovaných přístupových práv.

Životnost dat uložených v KDÚ je omezena životností formátu, resp. životností SW systému, kterým byla vytvořena. Zajištění "čitelnosti" v budoucnosti není tedy v okamžiku ukládání dat řešeno. Odkládá se na předem nespecifikovaný okamžik v budoucnosti, kdy bude třeba zajistit konverzi (migraci) formátu na nový formát, nebo jiné opatření, dostupné v budoucnosti.

Doporučujeme i v tomto případě zajistit omezení ukládaných formátů na rozumné minimum tak, aby náklady na úpravy formátů, nebo jiné řešení "čitelnosti" uložených dat zůstaly v rozumných mezích.

Dále navrhujeme pro takto uchovávaná data vytvořit katalog, kde bude udržován "globální" popis uložených dat.

Katalog bude udržovat/spravovat příslušný správce úložiště. Potřebné údaje musí získat od původce při iniciaci ukládání. Vybrané informace katalogu by podle potřeb mohly být dostupné i jiným osobám.

Schvalování iniciace ukládání nových dat (vznik nového logického segmentu) do KDU by měla být v kompetenci vedoucího IT kraje.

V závislosti na požadované době uložení dat v KDU je i zde třeba zajistit sledování životnosti použitých formátů. V pravidelných intervalech (například jednou za rok) správce katalogu vyhodnotí jednotlivé položky katalogu. Kromě aktualizace změněných údajů vyhodnotí životnost jednotlivých formátů (např. stavy: neohrožen, pravděpodobně ohrožen - zvýšená pozornost, ohrožen - nutno řešit, po době životnosti). O provedené kontrole vyhotoví strukturovaný zápis, který bude uložen pro audit.

Pokud je některý z formátů vyhodnocen jako ohrožený, zajistí potřebné kroky pro vyřešení tohoto stavu. V tomto případě bude třeba zjistit, jak bude ohrožený formát nahrazen a zajistit příslušné řešení.

1.1.1.1.11 Dimenzování KDÚ

Při dimenzování úložiště pro KDÚ vycházíme z odhadů množství dat uvedených v kapitole 4.1.1.1.7, ale tyto odhady dat jsou pouze orientační.

Z tohoto důvodu předpokládáme realizaci úložiště s možností postupného etapovitěho rozšiřování úložné kapacity.

Celková odhadovaná kapacita KDÚ na pět let je tedy **25,5 TB**. Vzhledem k potřebné rezervě a předem těžko odhadnutelným ročním přírůstkům doporučujeme kapacitu dimenzovat minimálně **na 30 TB**. Další kapacita by byla dokupována ročně dle potřeby.

Ostatní požadavky

1.1.1.1.12 Právní zakotvení KDR a KDÚ

Vzhledem k relativně nové problematice uchovávání digitálních archiválií nelze v současné době striktně stanovit právní postavení digitálních archivů na úrovni Kraje. Zákon č. 499/2004 Sb., v platném znění, pamatuje svými podmínkami a požadavky při zřizování těchto lokálních

archivů – archivů územních samosprávných celků - pouze na tzv. klasické archivy s klasickými, převážně listinnými archiváliemi. § 61 zákona stanoví, jaké podmínky takto akreditovaný archiv musí pro svou činnost splňovat a §§ 54 a 55 jeho oprávnění a působnost. Je na zvážení kraje a jeho orgánů, zdali tento akreditační proces postoupí a zařadí se tak do struktury oficiálních veřejných archivů působících v České republice.

Být akreditovaným archivem územního samosprávného celku především znamená, že:

- a) dohlíží na výkon spisové služby u svého zřizovatele a jím zřízených nebo založených organizací,
- b) pečuje o archiválie vzniklé z činnosti původců uvedených v předchozím písmenu,
- c) umožňuje nahlížet do archiválií u něj uložených,
- d) vede příslušnou evidenci archiválií podle tohoto zákona,
- e) vyhledává archiválie pro potřeby různých úřadů a osob, pořizuje z nich výpisy, opisy a kopie,
- f) provádí prověrku fyzického stavu archivních kulturních památek a národních kulturních památek, pokud jsou u něj uloženy,
- g) podílí se na sestavování celostátních tematických soupisů archiválií organizovaných ministerstvem,
- h) provádí konzervaci a restaurování archiválií,
- i) provádí inventuru archiválií vyhlášenou ministerstvem,
- j) provádí vědeckou, výzkumnou a vydavatelskou činnost v oblasti archivnictví, pomocných věd historických, dějin správy, regionálních dějin, dějin měst a obcí, historické vlastivědy a příbuzných oborů,
- k) zřizuje specializovanou knihovnu v rozsahu nezbytném pro plnění stanovených odborných a vědeckých úkolů,
- l) plní další odborné archivní úkoly na úseku archivnictví a výkonu spisové služby určené zřizovatelem.

Pro archiv tedy ze získané akreditace plynou určitá práva a povinnosti. Z těch práv je z pohledu zadání projektu nejdůležitější zákonná pravomoc vybírat a spravovat archiválie od svého zřizovatele a jeho zřízených nebo založených právnických osob a organizací. Pokud tento archiv získá akreditaci, vzniká jeho zřizovateli nárok na poskytování bezplatné odborné pomoci ze strany Národního archivu nebo příslušného státního oblastního archivu. Jedná se tedy o zákonem vymezené postavení archivu a získání určité vážnosti při poskytování archivních služeb pro veřejnost. Samozřejmě toto je vykoupeno povinnostmi, které by subjekt provozující pouhé úložiště dokumentů nemusel vykonávat.

Pokud se kraj rozhodne pro jiný statut tohoto archivu, jeho fungování a vztahy k badatelům a především původcům budou založeny pouze na smluvním základě a to buď tak, že bude tato instituce vystupovat v pozici uschovatele archiválií (§26 a následující zákona č. 499/2004 Sb.), případně nebude vůbec či zcela působit v oficiální rovině působnosti archivního zákona, ale např. jako shromažďovatel a poskytovatel nejen elektronických archiválií (případně jejich

„kopíí“), ale i dalších elektronických dokumentů, které ještě ani nemusí mít statut archiválie v pravém slova smyslu. V tomto případě je nutné upravit nejen vztahy k původcům, ale i případným zájemcům o bádání v těchto dokumentech. Při zřízení nebo založení takového subjektu (např. ve zřizovací či zakladatelské listině) by mělo být pamatováno i na to, zda subjekt nevykonává činnost, na kterou se vztahují předpisy o živnostenské nebo jiné podnikatelské aktivity a případně se podřídit i v nich uvedeným pravidlům.

Kromě smluvního základu mohou Zastupitelstvo Kraje, které má ve výhradní kompetenci zřizovat nebo zakládat organizační složky nebo právnické osoby a schvalovat jejich zřizovací či zakladatelské listiny nebo smlouvy, resp. Rada Kraje, která poté plní úkoly zřizovatele nebo zakladatele vůči zmíněným subjektům, přímým či nepřímým (faktickým) vlivem založit těmto podřízeným subjektům povinnost ukládat a uschovávat dokumenty v připravovaném krajském Technologickém centru v modulech KDS, KDR i KDÚ.

V případě zvolení modelu založeném na smluvním základě je však nutno vzít v úvahu možnou kolizi současných představ o různých typech ukládaných materiálů (typicky záznamy z kamerových systémů) s dalšími právními předpisy (zde zákon o ochraně osobních údajů – viz § 5 odst. 1 písm. e) – viz dále odstavec o ZOOÚ.

Lze předpokládat, že spolu se zákonem o Národním digitálním archivu budou aktualizována i ustanovení o archivech soukromoprávních původců a archivech územních samosprávných celků. Z diskutovaného návrhu by měly vzejít nejen podmínky, omezení, práva a povinnosti kladené na tyto digitální archivy, ale také vztahy k dalším subjektům a archivům.

Zákon o OOU (zákon o ochraně osobních údajů):

§5 odst. 1 písm e)

uchovávat osobní údaje pouze po dobu, která je nezbytná k účelu jejich zpracování. Po uplynutí této doby mohou být osobní údaje uchovávány pouze pro účely státní statistické služby, pro účely vědecké a pro účely archivnictví. Při použití pro tyto účely je třeba dbát práva na ochranu před neoprávněným zasahováním do soukromého a osobního života subjektu údajů a osobní údaje anonymizovat, jakmile je to možné,

§45 odst. 1

(1) Právnická osoba nebo fyzická osoba podnikající podle zvláštních předpisů se jako správce nebo zpracovatel dopustí správního deliktu tím, že při zpracování osobních údajů:

d) uchovává osobní údaje po dobu delší než nezbytnou k účelu zpracování [viz též § 5 odst. 1 písm. e)]

Vzhledem k tomu, že Kraj je správcem i zpracovatelem dokumentů uložených v KDR a je i správcem dat uložených v KDÚ, je třeba předem minimalizovat střet se zákonem o ochraně osobních údajů.

Právně nejkonfliktnější se jeví zamýšlené uchovávání záznamů z kamerových systémů v KDÚ.

Pokud Kraj chce tyto záznamy ve svém úložišti dlouhodobě uchovávat, doporučujeme získání akreditace archivu USC (Územního Samosprávného Celku) pro účely archivnictví dle výše uvedeného §5 ZOOÚ.

Uchování na dobu nezbytně nutnou resp., která je za takovou považována se rozumí max. 7 dní. V tomto smyslu hovoří i doporučení Výboru pro občanská a politická práva ohledně provozu kamerových systémů. Lze tedy důvodně předpokládat, že tato doba bude v dohledné době i zpracována do zákona.

1.1.1.1.1.13 Proč je dobré použít OAIS model

Model OAIS představuje primární a odbornou veřejností akceptovaný standard pro dlouhodobé uchování dokumentů. Je považován za směrnici „best practice“ v oblasti archivace (nejen) elektronických dokumentů. Jeho obliba vychází ze snadné pochopitelnosti, jejíž fungování lze snadno vysvětlit i laikům a také samozřejmě z vlastních kvalit tohoto modelu.

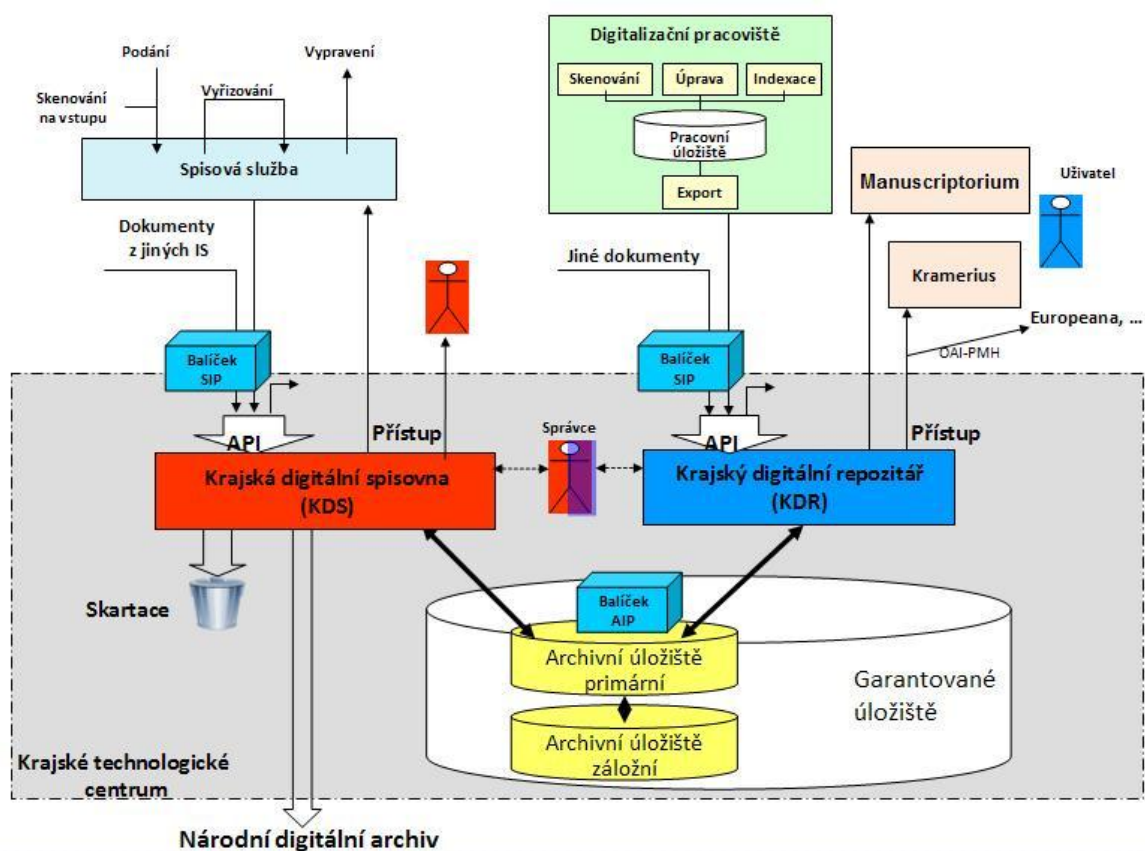
Znalost a použití tohoto modelu je celosvětově zakořeněna tak, že téměř žádné důvěryhodné úložiště se bez jeho aplikace neobejde. Z této všeobecné shody samozřejmě vyplývají určité výhody a jeho jednodušší použití:

- Vyzkoušené a prověřené ukládací politiky a postupy, které zajišťují bezpečné uchování informací a dokumentů v případě všech reálně zdůvodnitelných okolností
- Bezproblémová výměna, migrace dat mezi nezávislými úložišti splňující podmínky OAIS
- Snadná migrace dat najde uplatnění především při komunikaci mezi spisovnicemi, lokálními archivy a Národním digitálním archivem (NDA), kde se předpokládá model OAIS jako základ pro budování tohoto NDA
- Podobné požadavky na předávání dat jsou kladeny na dokumenty a jejich metadata i dle připravovaného Národního standardu
- V odborné veřejnosti známost a použití ustálených pojmů a paradigmat
- Srozumitelnost modelu nejen mezi odbornými pracovníky a knihovnicemi, ale i dalšími osobami – „být na druhé straně světa, ale mluvit stejnou řečí“
- Nabízí abstraktní a konceptuální řešení pro dlouhodobou archivaci dokumentů
- Univerzální model využitelný bez ohledu na velikost úložiště
- Nabízí adaptibilitu a rozšiřitelnost o vlastní funkční a informační potřeby při zachování OAIS compliance
- Lze poskytovat data jako autorizované kopie originálu s prokazatelnou a jednoduše vyhledatelnou historií o poskytovaných údajích
- Model, kde integrita, autenticita a prokazatelnost je zajišťována na velmi vysoké úrovni
- Repozitář je považován za důvěryhodný, pokud lze prokázat jeho schopnost plnit určité funkce, a pokud tyto funkce splňují minimální dohodnutá kritéria, jež by měly splňovat všechny důvěryhodné repozitáře. Rozhodující je, že naplňování těchto kritérií musí být prokazatelné, což znamená, že dosažení důvěryhodnosti je do velké míry závislé na auditu a certifikaci. Tzn., že každé dlouhodobé úložiště musí prokázat svoji důvěryhodnost, vitalitu a udržitelnost – existují různé standardy, certifikáty a audity,

které tyto kvality ověřují a úložiště certifikují. Většina je použitelná právě na model OAIS a i ty metody, které se proti použití pouze OAIS modelu vyhrazují, se téměř vždy stejně odkazují primárně na tento model.

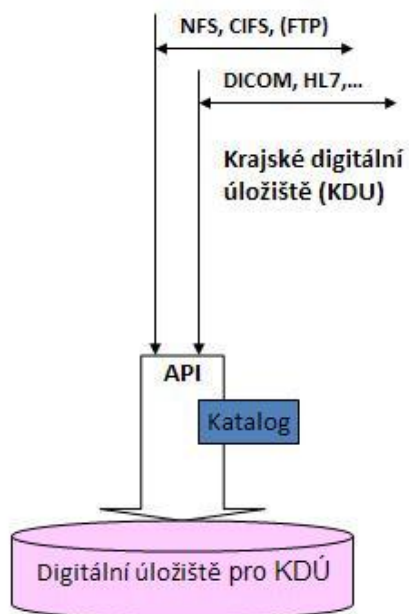
1.1.1.2 Softwarová architektura

Návrh architektury aplikací použitých při implementaci KDS a KDR vychází ze standardního modelu OAIS (viz předchozí kapitola). Struktura navržených komponent těchto dvou subsystémů je tedy podobná. Podstatné rozdíly jsou v konfiguraci funkcí, typu interních procesů, použití formátů a metadat, realizaci přístupu uživatelů a definici přístupových oprávnění.



Obrázek 4: Architektura KDS a KDR

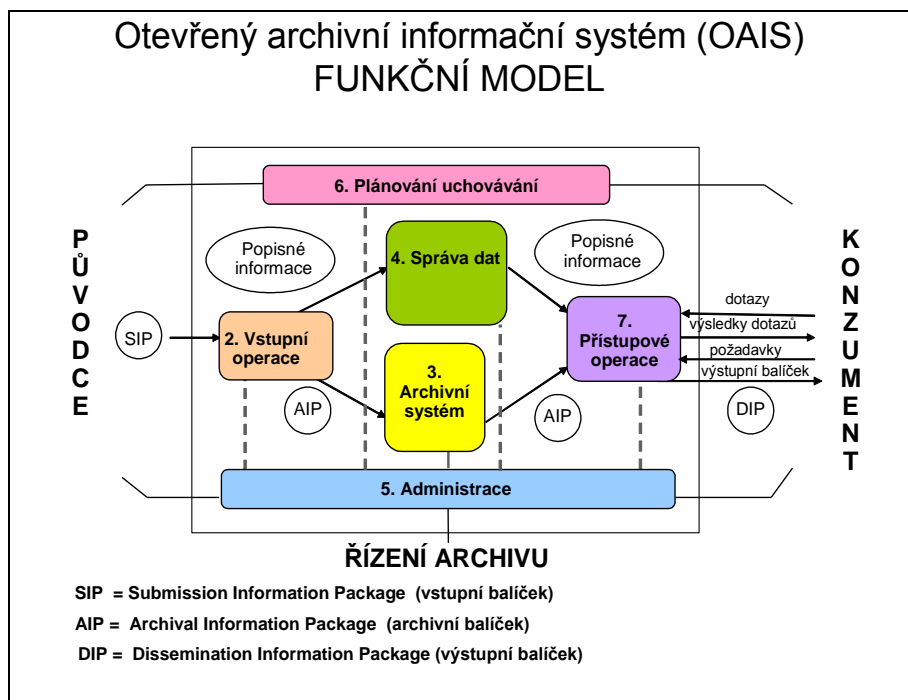
KDÚ je budováno jako systém pro zabezpečené uložení souborových dat a metadat bez jejich evidence a kontroly a s nižšími požadavky na zajištění dlouhodobého důvěryhodného uložení dokumentů při nižších nákladech na implementaci a provoz úložiště. Vzhledem k těmto cílům je softwarová architektura KDÚ podstatně jednodušší. Předpokládá se implementace i rozhraní API a UI využívající možností standardních souborových systémů.



Obrázek 5: Architektura KDÚ

Architektura OAIS

Zajištění dlouhodobého, důvěryhodného a i v budoucnosti čitelného uložení dokumentů a jejich srozumitelnosti i pro budoucí generace zabezpečuje v současnosti celosvětově nejčastěji přijímané řešení – vybudování otevřeného archivního informačního systému podle modelu OAIS. Funkční schéma modelu OAIS je uvedeno na následujícím obrázku:



Obrázek 6: Funkční model OAIS

Původce vytváří vstupní informační balíčky **SIP** tvořené vlastními dokumenty a jejich popisnými informacemi - metadaty.

V rámci **vstupních operací** mohou **administrátoři** archivu ještě doplňovat další potřebné **popisné informace** – metadata. OAIS kontroluje správnost těchto balíčků po stránce syntaktické a částečně i po stránce sémantické, vytváří z nich archivní informační balíčky **AIP** a ukládá je (dokumenty včetně jejich popisných informací - metadat) do speciálního archivního úložiště – **archivního systému**, který brání ztrátě informací v průběhu času metodami vícenásobného ukládání v různých lokalitách na různá média a nepřetržitými kontrolami obsahu balíčků AIP. Paralelně s tím ukládá OAIS samotné popisné informace z ukládaných AIP do provozního databázového systému OAIS - ve schématu označeného jako **správa dat**.

Konzument je subjekt oprávněný k přístupu k informacím z archivu. Konzument pomocí dotazů zjišťuje, jaké informace úložiště obsahuje (zjišťuje se pomocí **přístupových operací** nad správou dat) a OAIS mu zasílá výsledky těchto dotazů jako žádané odpovědi. Má-li konzument jasno v tom, jaká konkrétní data chce z archivu získat, pošle systému OAIS konkrétní požadavky a systém zajistí pomocí **přístupových operací** (nad správou dat a archivním systémem) vytvoření výstupního informačního balíčku **DIP** složeného z množiny archivních informačních balíčků AIP či množiny jejich částí.

norma ke stažení na: <http://public.ccsds.org/publications/archive/650x0b1.pdf>

Pro lepší pochopení principu funkčního modelu OAIS uvádíme podrobnější popis:

Prvním funkčním blokem modelu OAIS jsou "**Vstupní operace**", neboli "Příjem" (Ingest). Je to řada procesů zodpovědných za příjem informací od původců a jejich přípravu pro uložení do archivního úložiště. Přijímané informace se testují na kompletnost a neporušenost. Provádí se jejich transformace do tvaru vhodného pro jejich uložení a řízení v rámci archivního systému. Dále se provádí extrakce a tvorba (nebo pouze jedna z těchto akcí) popisných metadat

určených pro vyhledávání a zpřístupňování informací a přenos a ukládání předaných informací a k nim příslušných metadat do archivního systému. Stručně řečeno, funkce "Příjem" slouží jako vnější rozhraní OAIS k původcům, které řídí celý proces správy předávaných informací a jejich přípravy pro archivní uchování.

Druhým funkčním blokem je "**Archivní systém**" (Archival Storage). Je to část celého systému OAIS, která řídí dlouhodobé ukládání a údržbu elektronických dokumentů. Přesněji, tato funkce je zodpovědná za zajištění toho, že archivovaný obsah je uložen odpovídajícím způsobem a že bitový řetězec obsahující uchovávanou informaci bude dlouhodobě kompletní a čitelný. V archivním systému se implementuje řada bezpečnostních mechanismů, především periodické testování správnosti kontrolního součtu uložených informací a bezpečnostní politiky ku zmírnění následků katastrofických událostí. V případě zjištění poruchy nějakého archivního balíčku systém vytvoří okamžitě další bezpečnostní kopii z některého z identických správných balíčků, takže počet správných bezpečnostních kopií zůstává konstantní. Archivní systém podporuje požadavky konzumentů. Archivní systém ale nemá přímé externí rozhraní: interakce s ním zajišťují vysokoúrovňové služby systému OAIS.

Třetím funkčním blokem je "**Správa dat**" (Data Management). V rámci tohoto bloku se udržuje databáze popisných metadat identifikující a popisující archivované informace, která je pomocníkem při vyhledávání informací v archivu. Tento funkční blok řídí veškerá administrativní data podporující systém interních operací OAIS. Nejdůležitější funkce zahrnují databázové operace v souvislosti s dotazy, aktualizací archivu a dávkovými operacemi nad daty uloženými v archivu prostřednictvím metadat uložených v databázi.

Čtvrtou funkční částí je "**Plánování uchování**" (Preservation Planning). Tato služba odpovídá jak za plánování uchovací strategie OAIS, tak za doporučování příslušných revizí této strategie vycházející z podmínek okolí OAIS. Tato služba monitoruje změny vnějšího prostředí, které by mohly mít dopad na schopnost OAIS chránit a udržovat přístup k informacím v péči OAIS a rovněž tak sleduje inovace v paměťových a přístupových technologiích a v případném posunu oblasti zájmu nebo očekávání určené komunity. V rámci "Plánování uchování" se vytváří doporučení pro aktualizaci politik a procedur OAIS a pro přizpůsobení se těmto změnám. Tato funkce reprezentuje ochranu proti stále se vyvíjejícímu uživatelskému a technologickému okolí. Detekuje změny ovlivňující schopnost OAIS vyhovět své odpovědnosti, navrhuje strategie pro řešení těchto změn a podporuje implementaci těchto strategií v archivním systému.

Pátou funkcí OAIS jsou "**Přístupové operace**" (Access). Řídí procesy a služby, v rámci kterých konzumenti – a speciálně celá určená komunita – hledají, nacházejí, žádají a přijímají dodávku položek uložených v archivním úložišti OAIS. Typickou službou "Přístupu" je podpora konzumenta při vyhledávání položek, která předává žádost badatele "Správě dat" a výsledek dotazu prezentuje konzumentovi. Dále koordinuje získávání a dodávku žádaného obsahu konzumentovi. Předává příslušné požadavky "Archivnímu systému", získává žádané položky a provádí nutné transformace, které musí být provedeny před dodáním žádaných položek konzumentovi. Funkce "Přístup" je také zodpovědná za implementaci případných mechanismů přístupových práv svázaných s archivovaným obsahem. Funkce "Přístup" reprezentuje rozhraní OAIS ke konzumentovi a k určené komunitě. Provádí potřebné transformace včetně konverze archivovaných formátů na jeden či více formátů vhodných pro výstup, filtruje též pro výstup nepotřebná metadata.

Šestou a poslední funkcí OAIS je "**Administrace**" (Administration). Je odpovědná jak za řízení každodenních rutinních činností v rámci OAIS, tak za koordinaci aktivit v rámci OAIS. Komunikuje s původci (smlouvy o příjmu archivních dokumentů), s konzumenty (poskytuje zákaznickou podporu) a s vedením OAIS (implementace a údržba archivních politik a standardů). Funkce "Administrace" je také zodpovědná za supervizi archivních a přístupových operací, monitorování výkonnosti systému a za koordinaci aktualizace systému. "Administrace" je centrálním řídicím orgánem pro interní komunikaci (všechny popsané funkční celky OAIS) i externí komunikaci (původci, konzumenti, vedení OAIS).

Můžeme shrnout, že OAIS zahrnuje šest vysokoúrovňových funkčních částí, které, spojíme-li je dohromady, tvoří mechanismus pro dlouhodobé uchovávání informací, které též zpřístupňuje určené komunitě. Archivní systém založený na modelu OAIS implementuje každou z těchto služeb, přičemž formu této implementace nepředepisuje.

Balíčky AIP jsou základními datovými objekty v rámci systému vybudovaného dle modelu OAIS. Uchovávací metadata uložená jak v těle balíčku AIP, tak v Systému správy dat (Data Management), slouží pro podporu uchovávání a archivačních aktivit. Cílem pořizování uchovávacích metadat je podpoření pěti základních funkcí:

1. **životaschopnosti** (viability) – udržení digitálního objektu v bezpečí a neporušeného. Je nutné udržet integritu objektu pomocí objektivních kritérií. Využívat kontrolních součtů, zálohování, sledování změn apod.
2. **čitelnosti** (renderability) – udržení možnosti přehrání, otevření, zobrazení, spuštění a dalších možných přístupů k digitálnímu obsahu. Je nutné udržovat informace o potřebném prostředí pro využití digitálních zdrojů, jako je popis softwarového a hardwarového prostředí a prostředků pro práci s digitálními objekty. Dále je nutné udržovat popis struktury jednotlivých objektů a jejich vzájemné provázanosti.
3. **pochopitelnosti** (understandability) – zajištění pochopitelnosti obsahu pro budoucí použití. Na rozdíl od poskytování obsahu, kdy je důležitá jeho fyzická nebo syntaktická forma, pro pochopitelnost je důležité udržovat sémantickou integritu neboli význam a smysl uloženého obsahu.
4. **autentičnosti** (authenticity) – velkou obavou je udržení autentičnosti digitálního objektu a jeho citlivost na změny. Archivy potřebují dokumentovat jakoukoliv akci s digitálním objektem, aby budoucí uživatel mohl vyhodnotit jaké změny byly provedeny, včetně informací kdo a z jakého důvodu změnu prováděl. Je nutné zajistit, aby byly prováděny pouze kontrolovatelné a autorizované změny a ty byly dokumentovány (např. při provádění migrace formátu). Je nutné si uvědomit, že v případě použití certifikátů autenticity jsou tyto každou změnou porušeny a v současné době se případně používá re-autentizace a opětovné podepsání nové podoby digitálního objektu.
5. **identifikace** (identification) – zajištění jednoznačné identifikace digitálního objektu

Rozdíly mezi jednotlivými subsystemy

Pro jednotlivé typy dokumentů a dat v subsystémech KDS, KDR a KDÚ existují rozdílné požadavky na funkce i způsob uložení. Rozdíly mezi těmito daty s ohledem na požadovanou funkčnost systému představuje následující tabulka.

| Vlastnost | úřední dokumenty ve spisovně (KDS) | archivní dokumenty v repozitáři (KDR) | data v digitálním úložišti (KDÚ) |
|---|---|---|---|
| důvěryhodnost dokumentů či dat | Důvěryhodnost dokumentů je třeba vzhledem k právní prokazatelnosti prokazovat. U dokumentů získaných z datových schránek je důvěryhodnost zaručena tímto institutem, u dokumentů, který tímto kanálem neprošly je nutné zabezpečení elektronickým podpisem autora, a navíc zajištění periodickým přerazítkováním časovými razítky v rámci spisovny. | Důvěryhodnost zajištěna postačujícím způsobem politikami GÚ v rámci TC, právní prokazatelnost nemá u archivních dokumentů význam. | Důvěryhodnost dat zajištěna postačujícím způsobem politikami GÚ v rámci TC, v případě potřeby právní prokazatelnosti nezměněnosti uložených dat je možno data uložit na zařízení CAS / WORM. |
| přístupová práva k dokumentům či datům | Přístup musí být řízen systémem digitální spisovny - přístup k dokumentům musí být umožněn pouze pracovníkům původce podle jejich skupin (např. Pracovníci stavebního odboru) a rolí (uživatel, pracovník spisovny, administrátor, ...). Oprávnění je možno definovat ve vazbě na | Přístup musí být řízen systémem digitálního archivu, přístup musí být možno definovat podle nastavitelných politik na skupiny a role uživatelů. <input type="checkbox"/> některé dokumenty mohou být i obecně přístupné (veřejné) <input type="checkbox"/> nebo přístup k nim je řízen <ul style="list-style-type: none"> o autorskými právy | Přístup musí být řízen nastavením pravidel oprávnění technických prostředků úložiště (k jednotlivým adresářům filesystému) <input type="checkbox"/> přístup mají původci a jiné oprávněné osoby/orgány <input type="checkbox"/> je nutno přihlížet k ochraně osobních údajů |

| Vlastnost | úřední dokumenty ve spisovně (KDS) | archivní dokumenty v repozitáři (KDR) | data v digitálním úložišti (KDÚ) |
|-----------------------------------|--|--|--|
| | spisový plán. | <ul style="list-style-type: none"> ○ ochranou osobních údajů ○ komerčními hledisky (zhodnocení nákladů na pořízení dokumentů) | |
| metadatové standardy | <ul style="list-style-type: none"> <input type="checkbox"/> metadata stanovená Národním standardem pro elektronické systémy spisové služby (probíhá jeho schvalování) <input type="checkbox"/> použít stejný formát jako pro ukládání do Národního digitálního archivu (probíhá jeho schvalování) <input type="checkbox"/> povinná vazba na spisové plány původců | <ul style="list-style-type: none"> <input type="checkbox"/> standard metadat stanovený Národní knihovnou pro knihovní systémy (probíhá jeho příprava a schvalování) <input type="checkbox"/> možné vazby na číselníky stanovené Národní knihovnou (např. Celostátní databáze národních autorit vedená Národní knihovnou) <input type="checkbox"/> základní archivní metadata používaná při budování archivních fondů a sbírek <input type="checkbox"/> potenciálně může vzniknout požadavek i na využití standardů používaných v muzejnictví | <ul style="list-style-type: none"> <input type="checkbox"/> Metadatový standard nebude definován. Jednotlivá data-soubory nebudou opatřována metadaty. Globální metadatový popis bude dán strukturou katalogu KDÚ |
| způsob získávání metadat | Metadata předávána spolu s dokumenty na vstupu do spisovny. Metadata jsou u původce automaticky přebírána ze systémů spisové služby. | Metadata předávána spolu s dokumenty na vstupu do repozitáře. Podle typu ukládaných dokumentů se získávají většinou ruční indexací. Při digitalizaci je indexace jednou z činností digitalizačního pracoviště. U archivních fondů bude nutno metadata pořizovat až při ukládání do repozitáře – ruční práce odborných pracovníků. | Metadata v katalogu KDÚ bude spravovat správce katalogu KDÚ, potřebné údaje získá od původce při inicializaci vstupu. |
| požadavky na vyhledávání a | Specializované uživatelské rozhraní pro pracovníky původců, které musí | Specializované uživatelské rozhraní pro <ul style="list-style-type: none"> ○ administrátory repozitáře (kontrola | Technické rozhraní pro správu úložiště, jednoúčelové rozhraní pro |

| Vlastnost | úřední dokumenty ve spisovně (KDS) | archivní dokumenty v repozitáři (KDR) | data v digitálním úložišti (KDÚ) |
|---|---|---|---|
| komfort pro konzumenty | splňovat veškeré požadavky na interní procesy v úložišti (např. skartace), tak na vyhledávání i získávání dokumentů. Pro administrátory systému spisovny speciální rozhraní (kontrola vstupu dat, správa číselníku spisového plánu, apod.). | <p>vstupu dat, definice oprávnění, správa číselníků, apod.).</p> <ul style="list-style-type: none"> ○ odborné pracovníky (kontrola vstupu dat, doplnění metadat a pod.) ○ badatele ○ širokou veřejnost. To musí být jednoduché a naprosto srozumitelné bez jakéhokoliv školení | operátora KDÚ a další fyzické osoby s přidělenými přístupovými právy. |
| existence vnitřních řídicích procesů | Skartační řízení, zajištění důvěryhodnosti, zajištění čitelnosti. | Vnitřní skartace, rejstříkování, zpřístupňování, zajištění čitelnosti. | Žádné vnitřní řídicí procesy na vyšší úrovni, pouze v datovém úložišti periodické kontroly kontrolních součtů dat, jestliže to toto datové úložiště umožňuje. |
| spektrum technických formátů ukládaných dokumentů či dat | Úzké, definováno standardy. | Úzké, definováno standardy a metodikou repozitáře. | Neomezené. |
| způsob akvizice dokumentů či dat | "push" – sami smluvní původci ze své iniciativy zasílají dokumenty do spisovny. | "pull" – digitální archiv sám získává pro své fondy nové přírůstky vyhledáváním v různých zdrojích (fyzické a právnické osoby, jiné digitální archivy, knihovny) i "push" – výstupy z digitalizace . | "push" – sami původci ze své iniciativy zasílají data do KDÚ. |

| Vlastnost | úřední dokumenty ve spisovně (KDS) | archivní dokumenty v repozitáři (KDR) | data v digitálním úložišti (KDÚ) |
|---|--|--|---|
| doba uložení | Definovaná spisovým plánem max. cca 130 let. | Dlouhodobá – trvale. | Podle potřeby, dlouhodobost omezena životností formátů a SW. |
| zdroje dokumentů či dat | Omezený počet původců daný výčtem – Krajský úřad, úřady ORP měst a dalších obcí, zřizovaných organizací (v současnosti 1310 subjektů). Ukládání na základě smluvního vztahu (není legislativní podpora). | Neomezený počet původců včetně náhodných nebo neznámých zdrojů. Ukládání na základě smluvního vztahu. | Omezený počet původců daný výčtem (potřebami). Ukládání na základě smluvního vztahu. |
| frekvence přístupů k dokumentům či datům | Nízká, většinou při dohledávání nějakého staršího spisu. | Vysoká u dokumentů kulturního dědictví regionálního významu . | Nízká, přístupy pouze při haváriích zdrojových informačních systémů nebo speciální dotazy oprávněných orgánů. |
| počet současně pracujících konzumentů | Relativně malý počet. | V některých časových úsecích vysoký zájem jak o vyhledávání, tak pro získávání dokumentů kulturního dědictví regionálního významu. | Malý počet. |

Tabulka 11: Rozdíly mezi jednotlivými subsystémy

Softwarová architektura KDS a KDR

Subsystemy KDS a KDR založené na principech OAIS přistupují k ukládaným dokumentům a spisům jako k balíčkům, obsahujícím předmětná data a současně jejich metadata za účelem dlouhodobého uložení. Podle fáze jejich životního cyklu se jedná o vstupní (SIP), archivní (AIP) a výstupní (DIP) balíčky.

Rozhraní pro přístup k těmto systémům je specificky navrženo pro příjem a výdej balíčků v příslušném formátu definovaném na základě standardů. Vzhledem k zajištění bezpečnosti a konzistence uložených dat probíhá příjem dat do úložiště asynchronně v rámci procesu, který se skládá z několika kontrolních a transformačních procedur.

Systém digitálního archivu i digitální spisovny se skládá z těchto softwarových komponent:

▪ **Vstupní modul**

- Příjem dat

Zajišťuje komunikaci s původcem, autentizaci, autorizaci a uložení přijatých balíčků SIP do pracovního úložiště.

- Kontrola kvality vstupních dat (kontrola datové struktury, kontrola na obsah škodlivého kódu).

Kontroluje formální strukturu balíčků a přítomnost virů a jiného škodlivého obsahu balíčků. V rámci tohoto modulu je zřízena i tzv. karanténní zóna pro zajištění spolehlivosti kontrol.

- Řízení příjmu

Kontrola popisných a technických metadat, kontrola přípustnosti souborových formátů, kontrola struktury balíčku SIP a vzájemného provázání balíčků.

- Generování balíčků AIP

Automatické doplnění zejména technických metadat, konverze formátů metadat, možnost manuálního doplnění metadat, vstupní migrace formátů včetně generování náhledů pro prezentaci dat archivu v určeném formátu.

- Řízení ukládání

Zajišťuje konzistentní uložení metadat a obsahu archivních balíčků současně do archivního systému, systému správy dat a systému pro přístup.

▪ **Modul správy dat**

- Evidence číselníků

Zajišťuje ukládání a přístup k číselníkům používaným v rámci vstupní kontroly a vyhledávání. Jedná se zejména o tyto číselníky - původci, klasifikace, povolené

souborové formáty, kategorizace dokumentů podle kritérií přístupnosti, požadavků na zachování důvěryhodnosti, doby uložení.

- Evidence přijímaných a uložených balíčků.

Zajišťuje vedení a přístup ke katalogu uložených dokumentů včetně stavu příjmu a uložení.

- Evidence periodické obnovy časových razítek.

Zajišťuje evidenci historie obnovy časových razítek pro jednotlivé balíčky pro trvalé zajištění důvěryhodnosti uloženého obsahu.

- Evidence kontroly konzistence.

Uložení kontrolních součtů jednotlivých uložených balíčků AIP na aplikační úrovni pro účely periodické kontroly konzistence uloženého obsahu nezávisle na vlastnostech použitého archivního úložiště (CAS/NAS).

- Evidence procesů skartace a archivace.

Informace o stavu skartace a informace o stavu jednotlivých balíčků AIP zařazených do skartačního řízení.

▪ **Archivní systém**

- Zajišťuje vlastní důvěryhodné uložení obsahu balíčků AIP
- Je implementován primárně prostřednictvím technologie CAS nebo NAS

▪ **Modul administrace**

- Řízení procesu příjmu

Pro administrátora zajišťuje přehled o stavu příjmu balíčků SIP, umožňuje řešení problémů se strukturou a obsahem balíčků při příjmu.

- Řízení procesů migrace

Spouštění migrace souborových formátů v uložených balíčcích a přehled o provedených migracích.

- Řízení procesu časového razítkování.

Kontrola periodické obnovy časových razítek u uložených balíčků, případně i manuální spouštění obnovy razítek.

- Skartační řízení

Příprava návrhu a jeho schvalování, provedení skartace, případně exportu do jiného archivu v definovaném formátu.

- Správa kontroly konzistence

Přehled o průběhu ověřování kontrolních součtů a o nalezených problémech s uložením balíčků AIP.

- Správa číselníků.

Zajišťuje pro administrátory původce a archivu aktualizaci a čtení číselníků používaných v rámci vstupní kontroly a vyhledávání.

- Ukládání transakčních záznamů.

Pro účely auditu zaznamenává veškeré provedené operace nad uloženými balíčky (příjem, kontrola, transformace, ukládání, čtení). Zaznamenané záznamy jsou zároveň ukládány do úložiště ve formě AIP.

- Přístup k transakčním záznamům

Zobrazení transakčních záznamů pro účely auditu.

▪ **Přístupový modul**

- Zabezpečení přístupu a autentizace uživatelů.

Zajištění přístupu uživatelů k uloženým metadatům a dokumentům.

- Autorizace - omezení přístupů na základě klasifikace dokumentu, původce, uživatelských skupin a rolí uživatelů.

Modul povolí přístup ke čtení obsahu nebo metadat podle rolí přihlášeného uživatele a oprávnění příslušného balíčku.

- Vyhledání uložených balíčků na základě zvolených metadat.

- Zobrazení náhledů a distribuce uložených dokumentů ve formě DIP

Systém umožní výběr dokumentů a jejich zaslání oprávněnému uživateli ve standardizované podobě.

- Provádění transakčních záznamů o přístupu k jednotlivým uloženým balíčkům

- Programové rozhraní API na externí portál pro přístup

Systém eviduje veškeré přístupy k uloženým dokumentům a archivuje je.

Rozdíly ve funkčnosti KDS a KDR

▪ **Vstupní modul**

- Jiné standardy použité při kontrole a doplňování metadat.

V subsystému KDR se použijí standardy a metadata definovaná Národní knihovnou, případně další, která budou dohodnuta s původci archivních balíčků (knihovní sbírky, muzejní exponáty atd.).

V subsystému KDS se použije „Národní standard pro elektronické systémy spisové služby“ definovaný MV ČR a schéma pro předávání dokumentů a jejich metadat do archivu, které je definováno v rámci tohoto standardu.

- Obecně jiná konfigurace vstupních souborových formátů.

V subsystému KDS i KDR se využijí souborové formáty definované ve vyhlášce MV (vyhláška č. 191/2009 Sb.). V KDR navíc takové, které budou dohodnuty s původci archivních balíčků (knihovní sbírky, muzejní exponáty atd.).

- Odlišné nastavení použití archivního úložiště (volba způsobu uložení CAS/NAS). V subsystému KDS může být v některých případech (krátká skartační lhůta) použito úložiště typu NAS.

▪ **Modul správy dat a modul administrace**

- V KDS probíhá zákonné skartační řízení. V subsystému KDR se provádí pouze interní skartační řízení, t.zv. vnitřní skartace.
- V KDS probíhá opakovaná obnova časových razítek, v KDR probíhat nemusí.
- Odlišné nastavení přístupových oprávnění – u KDS jsou operace týkající se správy obsahu uložených dat delegována na pověřené správce původců (každý původce si spravuje svoji spisovnu), podobně přístup k uloženým dokumentům je pro uživatele původce. V KDR tyto operace typicky nevykonává přímo pracovník původce. Pro přístup k dokumentům v KDR je nutno mít nastavitelné politiky (obecně přístupné, omezení autorskými právy, osobních údajů a jiná omezení).

▪ **Přístupový modul**

- Datové balíčky KDR budou zpřístupněny prostřednictvím snadno ovladatelného přístupového modulu (t.zv. modulu zpřístupnění) větší množině uživatelů nebo veřejnosti.
- Datové balíčky KDS budou zpřístupněny pouze omezené množině autorizovaných uživatelů původce.

Softwarová architektura KDÚ

Subsystém KDÚ slouží k přímému, rychlému ukládání dat vybraných kategorií po dohodě s původci. Data jsou ukládána přímo ve formě souborů a povinně neobsahují popisná metadata v jednotně stanoveném formátu. Mimo zajištění spolehlivého uložení a zálohy obsahu datových souborů tento subsystém nezajišťuje žádné další obslužné operace sloužící k zajištění konzistence, důvěryhodnosti a přístupnosti obsahu.

Na úrovni Technologického centra kraje bude kromě vlastního úložiště nainstalován i příslušný software (např. archivní server pro ukládání zdravotní dokumentace ve formátu DICOM), který řídí jak komunikaci s archivující protistranou (původcem), tak i ukládání přenesených dat do fyzického úložiště. V tomto případě je komunikace po síti vedena na úrovni příslušného aplikačního protokolu a není využíván přímý přístup do souborového systému. Výhodou takového řešení může být snížení nákladů díky centralizaci archivačního subsystému a jeho obsluhy i možnost přímého využití specifických funkcí pro práci s archivovanými daty (vyhledávání, čtení, bezpečnost) informačním systémem původce.

Systém KDÚ je rozdělen na jednotlivé logické segmenty úložiště. Tyto logické segmenty jsou definovány v katalogu KDÚ a na jejich základě je vytvořena logická (adresářová) struktura úložiště. Pro každý segment je definován typ ukládaných dat, formát datových souborů, ukládací politika (pro řízení HSM), původce datových souborů a přístupová pravidla.

Systém digitálního archivu a digitální spisovny se skládá z těchto softwarových komponent:

▪ **Katalog KDÚ**

V katalogu jsou evidovány jednotlivé logické segmenty těmito parametry:

- Název logického segmentu a textový popis významu uložených dat
- Původce dat v logickém segmentu, jeho kontaktní osoby
- Definice typu ukládaných dat a formátu datových souborů v rámci logického segmentu
- Definice přístupového protokolu
- Způsob řízení životnosti dat v logickém segmentu
- Definice ukládací politiky požadovaného způsobu uložení s ohledem na rychlost přístupu (má vliv na konfiguraci HSM).
- Definice skupin uživatelů oprávněných k přístupu k souborům daného logického segmentu.
- Podrobný popis souborových formátů (dokumentace, standard) , kdo standard vydal a udržuje, kdo jiný standard ještě používá.
- Předpisy/normy podle kterých je třeba zajistit bezpečnost dat (osobní data, data chráněná autorským zákonem) v jednotlivých logických segmentech.
- Způsob kryptování, periodicitu obměny kryptovacích klíčů, dostupnost a způsob zajištění dostupnosti klíčů pro vybrané logické segmenty úložiště.

▪ **Úložiště**

- Obsahuje adresářovou strukturu vytvořenou na základě logických segmentů definovaných v katalogu KDÚ.
- Oprávnění přístupu k adresářové struktuře a uloženým datům jsou nastavena na základě definice oprávnění a kategorií podle údajů katalogu KDÚ.
- Předpokládá se realizace primárně s použitím technologie NAS

1.1.1.3 Technologická architektura

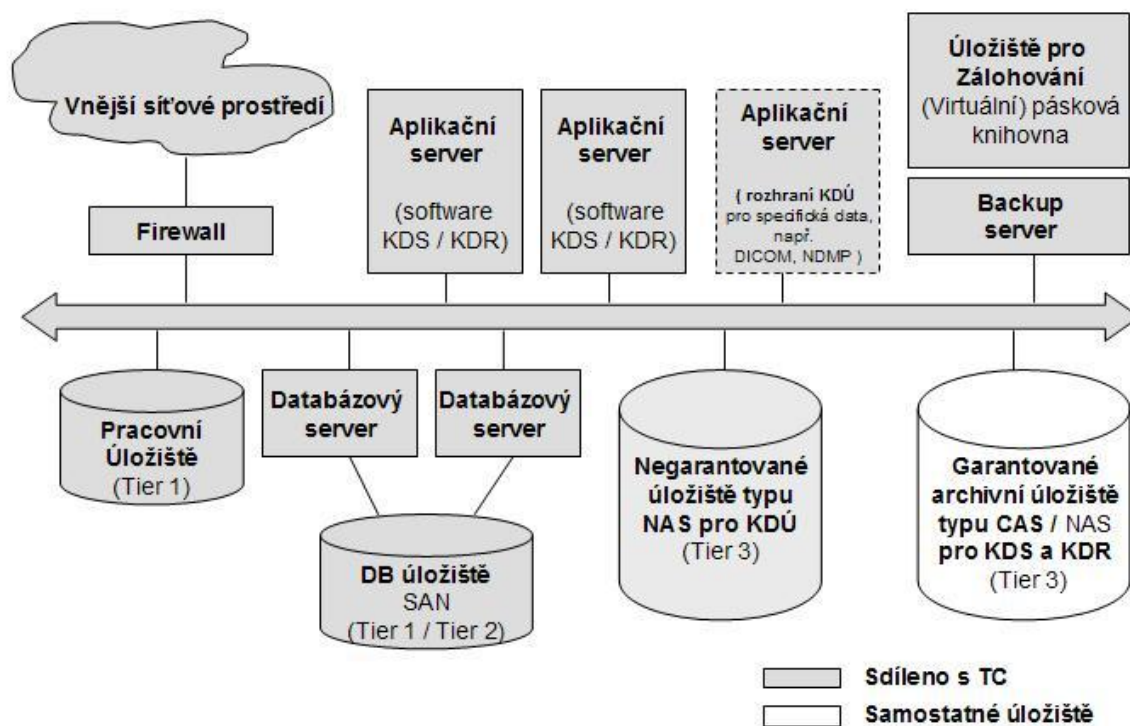
Systém bude využívat hardwarové a softwarové vybavení nasazené v rámci projektu Technologického centra Kraje (TC K).

V rámci tohoto projektu budou využity následující technologické části TC K:

- Aplikační servery pro obslužný SW subsystémů KDS a KDR
- Databázové servery využívané aplikacemi subsystému KDS a KDR
- Úložiště Tier 1 nebo 2 – pracovní prostory serverů a databáze subsystémů KDS a KDR

- Úložiště Tier 3 - technologie CAS nebo NAS pro subsystemy KDS a KDR a technologie NAS pro subsystem KDÚ. Předpokládáme zajištění replikací garantovaného archivního úložiště Tier 3 na dvě geografické lokality
- Systém zálohování TC K pro systémy a pracovní prostory serverů a databáze.
- Síťová infrastruktura TC K a zabezpečení přístupu z Internetu.
- Autentizace uživatelů.

Technologické schéma



Obrázek 7: Technologické schéma ukládání

Implementace subsystemů KDS a KDR

Digitální repozitář a digitální spisovna budou využívat specifické softwarové vybavení implementující příjem a správu dokumentů v intencích modelu OAIS. Tento software bude instalovaný na aplikačních serverech TCK, přičemž bude možné využít i virtualizace těchto serverů.

K obsluhým aplikacím KDS a KDR umístěným na aplikačních serverech bude možný vnější přístup pro jednotlivé původce dokumentů, kteří budou komunikovat prostřednictvím zabezpečeného kanálu (https) v rámci klientských aplikací a poskytovaných webových služeb.

Pro správu obslužných dat a metadat uložených balíčků subsystémů KDS a KDR bude využit databázový server a záložní databázový server podle potřeb těchto aplikací.

Pro účely důvěryhodného uložení balíčků AIP subsystémů KDS a KDR, obsahujících obsah dokumentů a jejich metadata, bude obslužnou aplikací použito přímo úložiště typu CAS nebo NAS, jehož obsah bude kompletně replikován v záložní lokalitě. Obslužná aplikace pracuje s daty uloženými v CAS nebo NAS i po jejich uložení (čtení obsahu po vyžádání, procesy pro zajištění důvěryhodnosti a konzistence).

Implementace subsystému KDÚ

Pro vybrané původce bude ve formě digitálního úložiště (KDÚ) zřízen zabezpečený přístup k jim přiděleným adresářům úložiště.

Pro tento účel bude využito negarantované úložiště NAS (Tier 3).

Na základě popisu jednotlivých typů logických segmentů v katalogu KDÚ budou pro jednotlivé adresáře úložiště definovány tyto vlastnosti:

- autorizace přístupu do úložiště
- uzamykání souborů (využití WORM vlastností NAS)

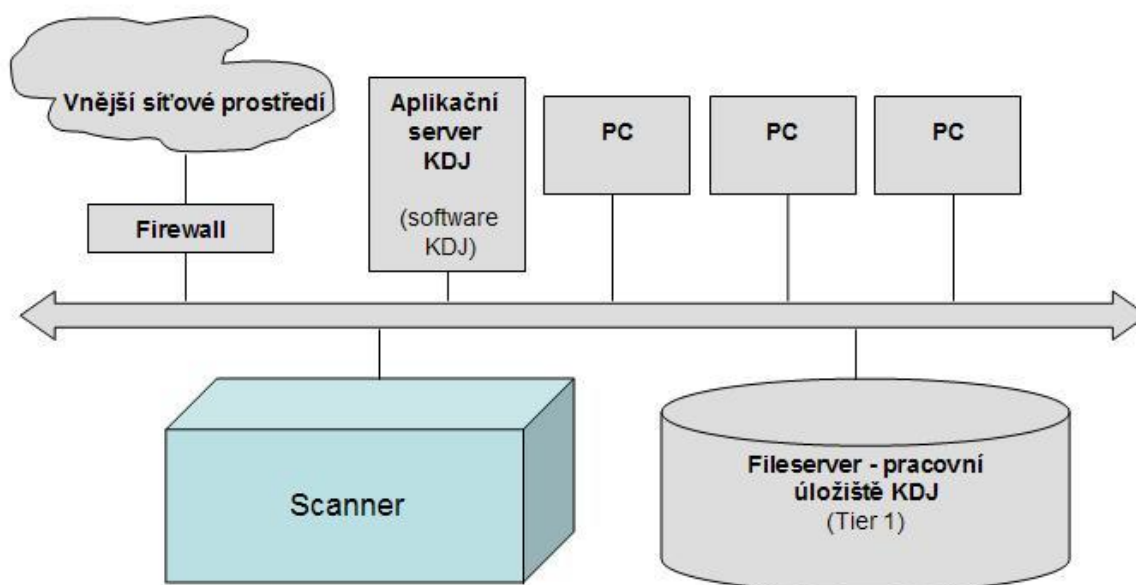
Implementace subsystému KDJ

KDJ bude implementována jako mobilní, čímž míníme, že bude instalována ve vhodných místnostech v bezprostřední blízkosti digitalizovaného fondu, ale počítá se s tím, že po digitalizaci tohoto fondu bude po pár měsících či rocích převezena do jiné lokality a tam opět instalována a využita k digitalizaci jiného fondu.

Základní umístění pracoviště KDJ bude v budově Krajského úřadu, kde bude KDJ umístěna na počátku projektu a pak i později dle potřeby. Součástí základního pracoviště bude i zamykací klimatizovaný mezisklad pro práci s historickými digitalizovanými fondy vybavený vysavačem pro odprášení a zamykacími vozíky na převoz dokumentů ke skeneru.

Principiální technologické schema KDJ je znázorněno na následujícím obrázku:

Technologické schéma mobilní KDJ



Obrázek 8: Technologické schéma mobilní KDJ

Sestava KDJ obsahuje tyto hlavní části

- kamerový knižní (univrzální) scanner formátu A2
- aplikační server
- datový fileservr o kapacitě cca 2TB jako pracovní datové úložiště
- 3 pracovní stanice – PC
- kabeláž
- připojení na silovou energii a rychlý internet

1.1.1.4 Uchovávací metoda (Digital preservation)

Strategie Digital preservation

"Digital preservation" je řízení ochrany digitálních informací v průběhu času. Ochranou digitálních informací je myšlena stálá a nepřetržitá pozornost věnovaná těmto informacím, nikoliv pouze ochraně médií, na nichž je digitální informace uložena. Nutnost stálého a

nepřetržitého věnování úsilí, času a peněz pro zajišťování technologických a organizačních postupů je hlavní překážkou v oblasti ochrany digitálních informací.

Ačkoliv jsme schopni přechít naše psané kulturní dědictví staré tisíce let, digitální informace vzniklé před pouhými deseti roky se nacházejí ve vážném nebezpečí, že budou nenávratně ztraceny.

V současné době existuje několik strategií pro dlouhodobé uchovávání digitálních informací, které je možné rozdělit na technologicky orientované strategie a informačně orientované strategie.

Technologicky orientované strategie uchovávání se zaměřují na zachování nebo udržování technologického prostředí pro digitální zdroje a jejich použití v původní podobě. Mezi tyto strategie patří:

- **údržba originálního prostředí** (uchovávání technologie „technologické muzeum“) – principem je údržba hardwarové a softwarové platformy pro podporu digitálního zdroje. Je svázána s obnovou médií, na kterém se digitální zdroj nachází. Jedná se o strategii nepraktickou a finančně náročnou. Nebezpečím je postupná ztráta znalostí obsluhy a fyzické zastarávání hardware bez možnosti obnovy. Tuto strategii lze ji použít na krátkodobé řešení uchování digitálního zdroje, než je možné použít strategii jinou. Nicméně např. pravidelná kontrola médií a jejich obnova je součástí správy uložených digitálních objektů i v jiných strategiích.
- **emulace** – principem je přizpůsobení prostředí datům; simuluje se původní hardwarové prostředí či původní operační systém na aktuální platformě. Dokumenty je tedy možné zobrazit v prostředí originálního software. Spolu s dokumenty se ukládají informace pro přizpůsobení prostředí. Tato strategie je vhodná např. pro uchování velkého množství dokumentů v omezeném formátu nebo pro spustitelné soubory.
- **virtualizace** – je do jisté míry kombinace emulace a migrace. Principem metody je převod dokumentu do spec jazyka (ve strojové řeči „Univerzálního virtuálního počítače“ UVC - Universal Virtual Computer) a jeho uchování spolu s programem, který umožní interpretaci dokumentu. Tuto metodu lze použít na velké množství digitálních zdrojů omezených formátů.

Informačně orientované strategie se zaměřují na čitelnost informací obsažených v digitálních zdrojích v budoucnosti, na budoucích technologiích. Mezi tyto strategie patří:

- **migrace na analogový formát** – principem je transformace digitálního objektu do analogové podoby. Příkladem je převedení snímání na mikrofilm. Ne všechny digitální objekty lze takto převádět a u mnohých hrozí ztráta vlastností. Strategie se používá například pro textové dokumenty.
- **migrace** – principem je přizpůsobení dat prostředí; elektronický dokument je převáděn z jedné HW či SW konfigurace na jinou, novější. V případě, že očekáváme zastarání formátu, ve kterém máme uložený obsah, provedeme dávkovou migraci. Uchovávaný obsah v takovém formátu převedeme hromadně do formátu nového.
- **encapsulace** – principem je zapouzdření digitálního objektu spolu s informacemi pro správnou interpretaci v digitálním objektu obsažených informací.

- **datová archeologie** – používá se při záchraně digitálních objektů na starých médiích a rekonstrukci uložených informací. Jako strategie je uváděna jako princip uložení digitálních objektů v původní formě, např. i z důvodu, že je nelze migrovat, a je pouze prováděna obnova média na kterém je digitální objekt uložen, s tím, že v budoucnu budou existovat nástroje schopné provést analýzu digitálního objektu a interpretovat jej.

Uchovávací metoda pro KDS a KDR

Vzhledem k potenciální možnosti přebírání zkušeností, standardů i pokynů pro dlouhodobou archivaci zmíněnou ve speciální kapitole navrhuje pro KDS a KDR použít pokud možno stejnou uchovávací metodu jako v projektech pro Národní digitální archiv a Národní digitální knihovnu.

S ohledem na pořizovací náklady, praktickou použitelnost pro řadu formátů a ověřenost praktickými zkušenostmi ve světě navrhuje použít primárně strategii migrace, spolu s omezením počtu povolených formátů. Snížením počtu povolených formátů, ať již daném legislativou (KDS) či pravidly repozitáře (KDR), se významným způsobem snižují potencionální náklady budoucí údržby (počet migrací, příprava migračních procedur).

Do budoucna však nelze vyloučit přechod na využití jiných uchovávacích strategií tak, jak bude postupovat jejich vývoj a výzkum.

Uchovávací metoda pro KDU

Pro KDU je uchovávací metodou podle předchozího rozdělení „údržba originálního prostředí“.

Aby se pokud možno zabránilo nutnosti použít v budoucnosti velmi drahou metodu datové archeologie, navrhuje dvě opatření:

- omezení množství povolených formátů na vstupu,
- evidence použitých formátů a pravidelná kontrola ohrožení formátů.

1.1.1.5 Dimenze konceptu budování a provozu KDS a KDR

Oba subsystémy, musí být budovány s perspektivou dlouhodobé existence. Po celou dobu musí být provozovány v rámci důvěryhodné instituce, kterou musí být TC Kraje.

Pro jejich vybudování doporučujeme postupovat podle mezinárodně uznávané metodiky PLATTER (PLanning Tool for Trusted Electronic Repositories), která zahrnuje všechny dimenze tohoto úkolu. Pamatuje tedy nejen na dimenzi technickou a technologickou.

V dalším popisu dimenzí používáme univerzální pojem "repozitář" (Repository) jak pro KDS, tak pro KDR, přičemž jak pro KDS, tak pro KDR musí být vypracovány konkrétní metodiky, které se v jednotlivých dimenzích odvodí z dále navrženého základu

Repozitář vybudovaný a průběžně udržovaný dle metodiky PLATTER dále zaručuje, že bude kdykoliv v budoucnosti úspěšně auditovatelný.

Právě monitoring všech dimenzí repozitáře zaručuje repozitáři jeho dlouhodobé a důvěryhodné fungování.

1. Finanční plán

- monitoring finančního plánu repozitáře a inicializace nutných akcí při jeho případných omezeních
- monitoring nových požadavků na finanční plán (ze všech ostatních oblastí strategického plánu) na tvorbu nového finančního plánu

2. Akviziční plán

- monitoring potencionálních původců a uzavírání smluv s nimi
- monitoring potřeb smluvních původců a inicializace potřebných akcí
- metodická podpora původců
- softwarová podpora původců
- parametrizace repozitáře podle potřeb původců
- údržba oblastí systémových číselníků závislých na původcích a metadat závislých na původcích

3. Plán řízení lidských zdrojů

- řízení repozitáře
- administrace repozitáře
- počty a kvalifikace pracovníků k zajištění dlouhodobého stabilního provozu
- minimální počty a kvalifikace pracovníků v době krizového řízení pro zajištění bezpečného uchování již uložených dokumentů

4. Plán zpřístupňování

- monitoring potencionálních konzumentů, začleňování do tříd, případné uzavírání smluv s nimi
- monitoring potřeb konzumentů a inicializace potřebných akcí
- metodická podpora konzumentů
- softwarová podpora konzumentů
- parametrizace repozitáře podle typů konzumentů
- údržba oblastí systémových číselníků závislých na typech konzumentů

5. Technologický plán

- údržba HW a SW repozitáře
- systémová administrace sítí
- systémová administrace SW repozitáře
- bezpečnostní audit repozitáře

6. Datový plán

- údržba číselníku datových formátů
- údržba specifikací datových zdrojů a přípustných formátů SIP podle původců
- programování procedur na převod datových formátů
- plánování potřebných převodů mezi datovými formáty pro účely (SIP -> AIP -> DIP)

7. Plán nástupnictví

- udržování kontaktů s jinými digitálními repozitáři s podobnou technologií pro případ vzájemného zabezpečení

8. Krizový plán

- Činnosti v případech předpovědatelných rizik (ekonomické otřesy, politické otřesy, ztráta mandátu, technické revoluce, přírodní katastrofy, ztráta uživatelů, vznik konkurenčního repozitáře, ztráta klíčových kompetentních pracovníků, průlom bezpečnosti)

9. Plán ochrany

- plánování migrací formátů souborů
- plánování migrací HW/SW repozitáře
- migrace všech souborů jednoho formátu v repozitáři na formát perspektivnější
- migrace HW/SW repozitáře

Metodiku PLATTER doporučujeme využít přiměřeně, s příslušnými omezeními i pro KDÚ.

1.1.2 Variantní návrhy technického řešení – HW/SW/data

Použije-li se pro archivní úložiště jakýkoliv typ technologie, musí systémy KDS / KDR spolu se systémem archivního úložiště každopádně zajistit tyto vlastnosti:

- trvalou garanci neměnnosti obsahu uložených archivních informačních balíčků AIP a jejich zajištění proti pozměnění obsahu třetí osobou
 - musí zajistit systém archivního úložiště
- zajištění ukládání a vyhledávání (čtení) archivních balíčků identifikovaných jménem, nikoliv jejich umístěním v úložišti
 - musí zajistit systém archivního úložiště
- automatické vytváření optimálně 4 identických kopií AIP (2 v jedné lokalitě, 2 v jiné lokalitě) a jejich AUTOMATICKOU periodickou kontrolu na kontrolní součet a AUTOMATICKÉ náhrady těch balíčků AIP, které byly zjištěny jako poškozené z těch identických kopií, které jako poškozené identifikovány nebyly
 - automatické vytváření více kopií musí zajistit systém archivního úložiště
 - automatickou periodickou kontrolu balíčků na kontrolní součet musí zajistit systém KDS / KDR
 - po odhalení chybového balíčku systémem KDS / KDR musí nastat komunikace mezi systémem KDS /KDR a systémem archivního úložiště a systém archivního úložiště musí poškozený archivní balíček nahradit jinou správnou kopií
- umožnění plánovaných kompletních periodických upgradů celého archivního úložiště v přelomových okamžicích celosvětového vývoje způsobů datové archivace, kdy se se veškeré AIP převedou do zcela nového archivního úložiště.
 - systém KDS / KDR překopíruje veškeré archivní balíčky do zcela nového úložiště

Variantní návrhy garantovaného archivního úložiště použitého pro KDS a KDR jsou následující:

Varianta 1 - úložiště typu CAS

CAS (Content-Addressable Storage) je nová kategorie automatizovaných síťových paměťových úložišť určených k dlouhodobému ukládání neměnného obsahu s podstatně delší dobou životnosti v porovnání s transakčními daty. Aplikace přistupují k úložišti typu CAS pomocí množiny API (aplikačních programových rozhraní). Primární přístupovou metodou je vyhrazené API, které podporuje práci s metadaty a využívá ostatních pokročilých vlastností úložiště CAS. Alternativně můžeme přistupovat k úložišti CAS více tradičními metodami, jako například pomocí protokolů NAS, FTP a HTTP, ovšem za cenu omezené funkcionality. Úložiště CAS poskytují on-line přístup k archivovaným datům s fixním obsahem, chrání jejich autenticitu jako WORM (Write-Once, Read-Many) a to za ceny TCO (Total Cost of Ownership) srovnatelné s páskami. Technologicky jsou systémy CAS většinou založeny na systémech RAIN (Redundant Array of Independent Nodes). Každá jednotka RAIN obsahuje procesor, software, a několik disků. Několik jednotek RAIN tvoří systém CAS. Celý CAS pak má své vlastní řízení, které zajišťuje redundantní definovatelné uložení na několik jednotek RAIN. Jednotky RAIN jsou většinou za provozu zaměnitelné ("hot swap"). Díky samokonfigurovatelnosti a vnitřním procesům kontroly konzistence dat a automatickým opravám případných výpadků se snižuje výrazným způsobem potřeba správy úložiště a tím celkové TCO. Disky jednotek RAIN se v případě nečinnosti zastavují, což významně přispívá k jejich dlouhodobé životnosti i k úspoře elektrické energie.

Varianta 2 - úložiště typu NAS

NAS (Network attached storage) je navrženo pro usnadnění spolupráce programů využívajících sdílení souborů. NAS zařízení se skládá z tzv. NAS hlavy nebo též tzv. NAS boxu, která představuje rozhraní mezi samotným diskovým zařízením a počítačovou sítí. NAS hlava řídí souborový systém na diskových zařízeních. Klient se připojuje na NAS hlavu, která má přidělenou IP adresu. NAS hlava převezme požadavek od klienta, zpracuje ho a získá data z diskového pole. Tato data poté NAS hlava pošle zpět klientovi. Tím je zajištěna plná transparentnost. Klientovi se celé NAS zařízení jeví jako jeden fyzický disk.

Zhodnocení:

Pro KDS / KDR je optimální úložiště typu CAS, protože je přímo pro tento účel zkonstruováno, ale je dražší variantou, nežli úložiště typu NAS. Systém typu NAS vyžaduje větší provozní podporu systémovými pracovníky. S ohledem na ekvivalentní vlastnosti obou variant je **doporučena varianta 2 – úložiště typu NAS**, protože se jedná o variantu s nižšími pořizovacími náklady.

1.1.3 Naplnění požadavků typizovaného projektu

Požadavky typizovaného projektu jsou předkládaným technickým řešením beze zbytku splněny a navíc je tímto řešením vyhověno požadavkům Karlovarského kraje na digitalizaci a ukládání v konkrétních podmínkách Kraje.

Navrhované řešení poskytuje navíc řadu konkrétních doporučení a know-how z oblasti digitalizace a ukládání.

1.2 Porovnání variant technologických řešení

V této kapitole je provedeno srovnání nabídek jednotlivých dodavatelů, jsou uvedeny výhody a nevýhody jednotlivých řešení a jsou analyzována technologická a bezpečnostní rizika.

1.2.1 Srovnání nabídek jednotlivých dodavatelů

Srovnání nabídek jednotlivých dodavatelů je uvedeno v následující tabulce. Nabídky byly poptávány v tomto provedení:

Popis řešení v této struktuře:

- použité technologie a rozhraní
- návrh architektury
- použité standardy, normy a zákonné vymezení
- způsob zajištění zaručující věrohodnost původu dokumentu, neporušitelnosti jeho obsahu a čitelnost dokumentu, a to včetně údajů prokazujících existenci dokumentu v digitální podobě v čase.
- návrh řešení HW části archivního (garantované) úložiště
- předběžný harmonogram implementace včetně prováděných činností

1.2.1.1 Krajská digitální spisovna

| Nabídka | Řešení | Výhody | Nevýhody | Rizika |
|---------------------|---|---|---|--|
| Nabídka č. 1 | Tessela SDB <i>LTP/DPS řešení</i> | Vyhovující použité technologie a rozhraní a návrh architektury. Návrh řešení HW části archivního (garantované) úložiště v souladu s TCK. Kapacitně dostatečně dimenzované řešení Nabídka předpokládá i společnou implementaci KDS a KDR. | Dlouhá doba pro implementaci Není reference v ČR | Neuvedeny použité standardy, normy a zákonné vymezení, speciálně ani soulad s 499/2004 Sb. Neuveden konkrétní původ "zajištění zaručující věrohodnost původu dokumentu, neporušitelnosti jeho obsahu ..." u aplikace pro KDS. V ČR dosud není implementace tohoto řešení. Vývoj některých funkcí může prodloužit a/nebo zdražit implementaci. Nutná lokalizace produktu. |
| Nabídka č. 2 | DESA <i>LTP/DPS řešení</i> | Vyhovující použité technologie a rozhraní a návrh architektury. Návrh řešení HW části archivního (garantované) úložiště v souladu s TCK. Nejrychlejší implementace Nabídka splňuje požadavky KDS Kapacitně dostatečně dimenzované řešení Řešení je vyvinuté na míru KDS Implementátor v ČR Řešení je vyvíjeno v ČR Nabídka předpokládá i společnou implementaci KDS a KDR | Není reference v ČR | V ČR dosud není implementace tohoto řešení. |
| Nabídka č. 3 | IBM Information Archive <i>HW řešení</i> | Návrh řešení HW části archivního (garantované) úložiště | Jedná jen o HW úložiště (Návrh řešení HW části archivního (garantované) úložiště) bez samotné aplikační nadstavby KDS. Nespĺňuje potřeby KDS. | Neuvedeny použité standardy, normy a zákonné vymezení, speciálně ani soulad s 499/2004 Sb. Neuveden harmonogram implementace. Nedostačující řešení KDS. V ČR dosud není implementace tohoto řešení pro potřeby KDS. |

| | | | | |
|----------------------------|--|---|--|--|
| <p>Nabídka č. 4</p> | <p>STA - O2 důvěryhodné úložiště <i>LTP/DPS řešení</i></p> | <p>Vyhovující použité technologie a rozhraní a návrh architektury. Návrh řešení HW části archivního (garantované) úložiště v souladu s TCK. Kapacitně dostatečně dimenzované řešení Nabídka předpokládá i společnou implementaci KDS a KDR.</p> | <p>Není reference v ČR</p> | <p>Neuveden harmonogram implementace. V ČR dosud není implementace tohoto řešení.</p> |
| <p>Nabídka č. 5</p> | <p>OpenText <i>DMS řešení</i></p> | | <p>Nejedná je o systém pro potřeby LTP-DPS. Řešení nesplňuje základní podmínky, např. soulad s OAIS.</p> | <p>Nabídka od dodavatele nedorazila v požadované struktuře a obsahu. Neuveden popis aplikace KDS/R. Neuvedeny použité standardy, normy a zákonné vymezení, speciálně ani soulad s 499/2004 Sb. Riziko značných víceprací pro dosažení shody s požadavky, kladenými na KDS. Navržené řešení není stavěno na dlouhodobé ukládání. Neuveden harmonogram implementace. V ČR dosud není implementace tohoto řešení.</p> |
| <p>Nabídka č. 6</p> | <p>FileNet <i>DMS řešení</i></p> | | <p>Nejedná je o systém pro potřeby LTP-DPS. Řešení nesplňuje základní podmínky, např. soulad s OAIS.</p> | <p>Nabídka od dodavatele nedorazila v v požadované struktuře a obsahu. Neuveden popis aplikace KDS/R. Neuvedeny varianty KDS/KDR/KDS+KDR. Neuvedeny použité standardy, normy a zákonné vymezení, speciálně ani soulad s 499/2004 Sb. Riziko značných víceprací pro dosažení shody s požadavky, kladenými na KDS. Navržené řešení není stavěno na</p> |

| | | | | |
|--|--|--|--|--|
| | | | | <p>dlouhodobé ukládání. Neuveden harmonogram implementace. V ČR dosud není implementace tohoto řešení.</p> |
|--|--|--|--|--|

Tabulka 12: Přehled nabídek pro KDS

1.2.1.2 Krajský digitální repozitář

| Nabídka | Řešení | Výhody | Nevýhody | Rizika |
|---------------------|--|---|--|---|
| Nabídka č. 1 | TesselaSDB+IDEA@ALM <i>LTP/DPS řešení</i> | Nabídka splňuje požadavky KDR Kapacitně dostatečně dimenzované řešení Nabídka předpokládá i společnou implementaci KDS a KDR | Délka implementace (8 měs) Viz. KDS | Závislost na OS a platformě jednoho výrobce Viz. KDS |
| Nabídka č. 2 | DESA <i>LTP/DPS řešení</i> | Rychlá implementace (do 3 měs) Nabídka splňuje požadavky KDR Kapacitně dostatečně dimenzované řešení Řešení je vyvíjeno v ČR Nabídka předpokládá i společnou implementaci KDS a KDR | Viz. KDS | Viz. KDS |
| Nabídka č. 3 | IBM Information Archive <i>HW řešení</i> | Viz. KDS | Viz. KDS | Viz. KDS |
| Nabídka č. 4 | Rosetta/ Digitoool <i>LPT řešení</i> | Viz. rizika | Viz. rizika | Nabídka od dodavatele nedorazila v požadované struktuře a obsahu. Nabídka neobsahovala žádný popis řešení, jen cenovou kalkulaci KDR. V ČR dosud není implementace tohoto řešení. |

| Nabídka | Řešení | Výhody | Nevýhody | Rizika |
|--------------|--|----------|----------|----------|
| Nabídka č. 5 | STA - O2 důvěryhodné úložiště <i>LTP/DPS řešení</i> | Viz. KDS | Viz. KDS | Viz. KDS |
| Nabídka č. 6 | OpenText <i>DMS řešení</i> | Viz. KDS | Viz. KDS | Viz. KDS |
| Nabídka č. 7 | FileNet <i>DMS řešení</i> | Viz. KDS | Viz. KDS | Viz. KDS |

Tabulka 13: Přehled nabídek pro KDR

1.2.1.3 Digitalizované a uložené dokumenty

| | | Jednotka | Nabídka 1 | Nabídka 2 | Nabídka 3 | Předpokl. cena do studie |
|--------------------------|--|--------------------|-----------|-----------|-----------|--------------------------|
| Knihy a periodika | Rozměr A5 až A3, 600dpi plnobarevně | Kč/strana bez DPH | 12,90* | 5,10* | 6,05* | 8,00 |
| | Rozměr A5 až A3, 300dpi ve stupních šedi | Kč/ strana bez DPH | 9,90* | 4,00* | 4,90* | 6,00 |
| Dokumenty | Rozměr A3 až A4, 600dpi stupně šedi | Kč/ strana bez DPH | 2,90 | 3,65 | 3,25 | 3,00 |
| | Rozměr A2-A0, 300dpi plnobarevně nebo ve stupních šedi | Kč/ strana bez DPH | 31,20 | 47,30 | 31,20 | 32,00 |

Tabulka 14: Ceny za digitalizaci dokumentů

* Uvedené ceny nezahrnují některé další poplatky za zpracování OCR, manipulaci, práci s poškozenými tisky, přípravu metadat atd. Tyto služby všechny oslovené firmy nabízejí, cení se buď procentem z ceny, nebo pevnou částkou za stránku. Typicky je za tyto služby účtována přírážka cca 0,50-1,5Kč/stranu za automatické OCR a 3-5 Kč/stranu za zpracování metadat.

Pro další potřebu stanovení nákladů ve studii proveditelnosti byly navrženy ceny, které jsou uvedené v pravém sloupci. Cena za digitalizaci knih nezahrnuje automatické OCR a zpracování metadat.

Pokud byly v nabídkách uvedeny ceny za různé množství kategorií, byly ceny počítány pro množství 1 milion stran.

Obecně lze říci, že nabídky potvrdily proveditelnost digitalizace fondu kraje formou služby v rámci předpokládaného rozpočtu.

1.2.1.4 Krajská digitalizační jednotka

| Krajská digitalizační jednotka | Investice v Kč včetně DPH |
|--|------------------------------|
| Velkoplošný skener do formátu A0 s knižní kolébkou (ProServ ScannTech, Planscan, CONTEX, ELSYST ENGINEERING ATLAS) | 1 500 000 Kč až 2 500 000 Kč |
| Kamerový knižní (univerzální) skener do formátu A2 (BookEye, Zeutschel) | 650 000 Kč až 1 100 000 Kč |
| Infrastruktura pro zpracování skenů (SW na úpravy skenů, standardní aplikační server, standardní datový fileservr s kapacitou do 2 TB, 3 kusy standardních PC) | 500 000 Kč až 600 000 Kč |

Tabulka 15: Ceny za KDJ

Oslovené firmy většinou nabídly ceníkové ceny, které byly vzájemně srovnatelné. Nabídka produktů v této oblasti není příliš rozsáhlá, proto se většina nabízených modelů opakovala ve více nabídkách. Cena pravidelné údržby se pohybuje většinou okolo 10% z pořizovací ceny.

V kategorii „Kamerový knižní (univerzální) skener do formátu A2“ jsou nabídky podle katalogových hodnot víceméně srovnatelné. Rozdíly v cenách jsou dány i tím, že součástí některých cen je i implementace a zaškolení obsluhy. Rozdíly v kvalitě se však mohou projevit až ve chvíli testování skeneru v reálném provozu.

1.2.2 Výhody a nevýhody jednotlivých řešení

Porovnání výhod a nevýhod jednotlivých řešení je uvedeno v kapitole 1.2.1.

1.2.3 Analýza technických a bezpečnostních rizik

Technická a bezpečnostní rizika jednotlivých řešení jsou uvedena v kapitole 1.2.1.

Jediné riziko „V ČR dosud není implementace tohoto systému.“, které je společné pro všechny nabídky, odpovídá aktuální situaci digitálních spisoven v ČR, nejedná se však o riziko závažného charakteru.

1.3 Doporučení a upřesnění pro účely zadávací dokumentace a realizační projektové dokumentace

Vlastní projekt Digitalizace a ukládání bude rozdělen do dvou veřejných zakázek, ty budou vzhledem k výši zakázky a dle Zákona o veřejných zakázkách řešeny jako nadlimitní veřejné zakázky v otevřeném řízení.

1.3.1 Specifikace zadání technického řešení

V rámci projektu se předpokládá realizace dvou nadlimitních veřejných zakázek, které budou členěny do částí, v následujícím rozsahu:

1) Veřejná zakázka na digitalizační služby, dodávku vybavení KDJ

- a) Služba digitalizace knižního fondu
- b) Služba digitalizace dokumentů
- c) Dodávka skenerů
- d) Dodávka skenovacího software
- e) Dodávka pracovních stanic
- f) Opční právo na další rozšíření kapacity diskového pole na kapacitu nutnou pro fázi provozu

2) Veřejná zakázka na vytvoření a údržbu krajských dlouhodobých úložišť

- a) Dodávka a implementace KDS
- b) Dodávka a implementace KDR
- c) Dodávka a implementace KDU

Důvodem rozdělení veřejných zakázek je vzájemná technologická nezávislost.

Rozsah předmětu veřejné zakázky je dán technickým řešením projektu (viz kapitola 1.1).

Zadávací dokumentace veřejné zakázky musí obsahovat požadavky specifikace technického řešení, která je popsána v kapitole 1 a musí splňovat i podmínky SF EU – výzvy IOP č. 08.

Opční právo u části 1f) je využito z důvodu odkladu nákupu kapacity na konec realizační fáze – pořizovat plnou kapacitu diskového úložného prostoru hned na začátku realizační fáze je zbytečné a navíc ekonomicky nevýhodné.

Realizační dokumentace bude vypracována na základě konkrétní dodávky vyhlášené veřejné zakázky na vytvoření a údržbu krajských dlouhodobých úložišť.

Technickou podporu bude zajišťovat dodavatel řešení. Proto musí být požadavkem na dodavatele dostatečné technické i personální zázemí, zkušenosti s technickou podporou a prokázaná schopnost zajistit podporu provoz řešení dlouhodobě a kvalitně.

Pro zajištění publicity projektu bude dle Zákona o veřejných zakázkách vybrán dodavatel na v investiční fázi projektu.

1.3.2 Požadavky na implementaci, školení a technickou podporu

Vlastní implementace musí být zajištěna kvalifikovaným realizačním týmem dodavatele, který bude doplněn realizačním týmem zadavatele, který zajistí odpovídající součinnost zadavatele. Implementace bude probíhat dle odsouhlaseného harmonogramu. Součástí implementace bude i školení odpovídajícího počtu tzv. „klíčových uživatelů“ a pracovníků zadavatele pověřených administrací a údržbou implementovaného řešení.

HW a SW komponenty, klíčové pro zajištění chodu KDS, KDR a KDU budou pokryty smlouvou o podpoře, která bude obsahovat právo na nové verze i technickou podporu výrobce (nebo výrobcem certifikované servisní organizace) a to nejméně po dobu udržitelnosti projektu. V případě KDS bude smlouva zahrnovat zajištění souladu s platnou legislativou po celou dobu platnosti.

1.4 Provozní zajištění projektu

Provoz jednotlivých komponent projektu bude zajišťovat Karlovarský kraj, buď vlastními zaměstnanci, nebo s pomocí externí firmy.

Požadavky na provozní zajištění jsou definované v oblastech energetických a materiálových toků, záruky a servisu, údržby, životnosti a provozní náročnosti.

1.4.1 Potřebné energetické a materiálové toky

Energetické toky zahrnují pouze napájení digitalizační jednotky, kde je spotřeba srovnatelná s jinou běžnou kancelářskou technikou tj. zanedbatelná. Jiné energetické toky nejsou pro tento projekt uvažovány, zbytek systému bude provozován v prostředí Technologického centra.

Významným materiálovým tokem bude předávání podkladů k digitalizaci dodavateli služby – prodiskutovat jak bude probíhat – navrhujeme předávání cca 1x měsíčně.

Ostatní materiálové toky jsou zanedbatelné, jedná se pouze o běžný administrativní spotřební materiál.

1.4.2 Záruky a servis

Záruční doba bude sjednána minimálně na dobu 24 měsíců ode dne předání předmětu k užívání.

Dodavatel ve své nabídce specifikuje:

- proces reklamace,
- reakční doby,
- požadavky na součinnost,
- další práva a povinnosti dodavatele i zadavatele.

1.4.3 Údržba a nákladnost oprav

Údržba a odstranění nedostatků budou prováděny v rámci záruční doby. Na všechny HW komponenty, pořizované v rámci tohoto projektu, bude uzavřena servisní smlouva, aby byl zajištěn odborný servis těchto zařízení po dobu udržitelnosti projektu.

1.4.4 Údaje o životnostech jednotlivých zařízení

KDS, KDR a KDU budou provozovány v prostředí TC K, životnost jednotlivých zařízení proto není v rámci projektu relevantní, je dána životností HW technologií, na kterém jsou úložiště provozována. Projekt předpokládá, že životnost těchto technologií přesahuje dobu udržitelnosti projektu.

V případě pracovních stanic je jejich technologická životnost také delší, než je doba udržitelnosti projektu. Je však možné, že z důvodů morálního zastarávání nebo dodržení standardů bude nutné je vyměnit za novější ještě před uplynutím doby udržitelnosti.

1.4.5 Údaje o provozním zajištění SW a datových komponent

Provoz bude zajištěn výše navrhovaným způsobem servisu. Délka trvání záruční doby je nastavena na standardních 24 měsíců. Činnosti nad rámec záruky budou pokryty provozní smlouvou mezi Karlovarským krajem a dodavatelem řešení. Veškeré náklady budou plně hrazeny krajem.

SW komponenty, klíčové pro zajištění chodu KDS, KDR a KDU budou pokryty smlouvou o podpoře, která bude obsahovat právo na nové verze i technickou podporu výrobce (nebo výrobcem certifikované servisní organizace) a to nejméně po dobu udržitelnosti projektu. V případě KDS bude smlouva zahrnovat zajištění souladu s platnou legislativou po celou dobu platnosti.

Ostatní SW komponenty budou pořízeny v takových verzích, aby je výrobce podporoval bezplatně minimálně po dobu udržitelnosti projektu. Podporou se zde rozumí minimálně poskytování bezpečnostních a funkčních oprav (patchů) a přístup k databázi známých řešených problémů.

1.4.6 Změny v provozní náročnosti vlivem opotřebení

U KDS, KDR a KDU se jedná o softwarová řešení, u kterých nedochází k opotřebení.

HW technologie, pořizované v rámci tohoto projektu, mají víceméně konstantní provozní náročnost po dobu životnosti, která přesahuje dobu udržitelnosti. Provozním opotřebením se nezvyšuje se spotřeba elektrické energie ani náročnost údržby či profylaxe.