



# Studie regionálního konceptu digitalizace a elektronické archivace kraje Vysočina

ICZ a.s.

Vysočina  
Žižkova 57/1882  
587 33 Jihlava

Copyright © 2009 ICZ a.s.

Žádná část tohoto dokumentu nesmí být kopírována žádným způsobem bez písemného souhlasu majitelů autorských práv. Autorská a jiná díla odvozená z tohoto díla podléhají ochraně autorských práv vlastníků.

**Historie verzí**

Číslo verze	Datum verze	Vypracoval	Popis	Jméno souboru
0.7	2. 7. 2009	ICZ, a.s.	Draft	Studie_Vysocina_Draft_090702.pdf
1.0	10.7.2009	ICZ, a.s.		Studie_Vysocina_V1_090710.pdf
1.1	23.7.2009	ICZ, a.s.	Úpravy dle připomínek	Studie_Vysocina_V1-1_090723.pdf
1.2	13.10.2009	ICZ, a.s.	Terminologické úpravy	Studie_Vysocina_V1-2_091013.pdf

# Obsah

<b>OBSAH</b> .....	<b>3</b>
<b>1 MANAŽERSKÉ SHRUTÍ</b> .....	<b>5</b>
1.1 Předmět zakázky.....	5
1.2 Nástin řešení.....	6
<b>2 RÁMEC ŘEŠENÍ</b> .....	<b>8</b>
2.1 Vymezení důležitých pojmů:.....	8
2.2 Problematika digitalizace.....	9
2.3 Problematika ukládání a archivace.....	14
2.3.1 Dlouhodobá archivace dokumentů.....	15
2.3.2 Bezpečné dlouhodobé ukládání dat.....	18
<b>3 ZDROJE DAT (SOUČASNÝ STAV PROBLEMATIKY)</b> .....	<b>18</b>
3.1 Krajská digitální spisovna.....	18
3.1.1 Původci dokumentů.....	18
3.1.2 Zdrojové systémy.....	19
3.1.3 Odhad množství dokumentů.....	19
3.1.4 Identifikace a kategorizace dokumentů, metadata.....	20
3.1.5 Formáty elektronických dokumentů.....	21
3.2 Krajský digitální repozitář.....	21
3.2.1 Původci dokumentů.....	21
3.2.2 Zdrojové systémy.....	22
3.2.3 Odhad množství dokumentů.....	22
3.2.4 Identifikace a kategorizace dokumentů, metadata.....	23
3.2.5 Formáty elektronických dokumentů.....	23
3.3 Krajské digitální úložiště.....	24
3.3.1 Odhad množství dat.....	24
3.3.2 Identifikace a kategorizace dat, metadata.....	25
3.3.3 Formáty dat.....	26
3.4 Souhrnný odhad velikosti fyzického úložiště.....	26
<b>4 POŽADAVKY NA BUDOUCÍ SYSTÉM</b> .....	<b>27</b>
4.1 Architektura OAIS.....	27
4.2 Rozdíly mezi jednotlivými subsystemy.....	30
4.3 Krajská digitální spisovna (KDS).....	35
4.4 Krajský digitální repozitář (KDR).....	35
4.5 Krajské digitální úložiště (KDÚ).....	36
4.6 Ostatní požadavky.....	37
4.6.1 Právní zakotvení KDR a KDÚ.....	37
4.6.2 Proč je dobré použít OAIS model.....	39
<b>5 GLOBÁLNÍ KONCEPCE BUDOUCÍHO SYSTÉMU</b> .....	<b>40</b>
5.1 Softwarová architektura.....	40

5.1.1	Softwarová architektura KDS a KDR .....	41
5.1.2	Rozdíly ve funkčnosti KDS a KDR .....	43
5.1.3	Softwarová architektura KDÚ.....	44
5.2	Technologická architektura .....	45
5.2.1	Implementace subsystémů KDS a KDR .....	45
5.2.2	Implementace subsystému KDÚ.....	46
5.3	Uchovávací metoda (Digital preservation) .....	46
5.3.1	Strategie Digital preservation.....	46
5.3.2	Uchovávací metoda pro KDS a KDR: .....	47
5.3.3	Uchovávací metoda pro KDÚ: .....	48
<b>6</b>	<b>ZÁVĚRY A DOPORUČENÍ.....</b>	<b>48</b>
6.1	Dimenze konceptu budování a provozu KDS a KDR .....	48
6.2	Návrh etapizace a příslušných pilotních projektů .....	49
6.2.1	První etapa .....	50
6.2.2	Druhá etapa.....	50
6.2.3	Třetí etapa .....	50
6.2.4	Čtvrtá etapa .....	50
6.2.5	Rutinní provoz.....	51
6.3	Rizika projektu .....	51
<b>7</b>	<b>PŘÍLOHY .....</b>	<b>54</b>
7.1	Spisový plán Krajského úřadu Vysočina – vybrané položky .....	54
7.2	ZKRATKY A POJMY .....	56
7.2.1	Zkratky .....	56
7.2.2	Pojmy .....	59

<b>Název veřejné zakázky malého rozsahu:</b>	<b>„Studie regionálního konceptu digitalizace a elektronické archivace kraje Vysočina“</b>
<b>Zadavatel</b>	
<b>Název:</b>	Vysočina
<b>IČ :</b>	70890749
<b>Adresa sídla:</b>	Žižkova 57/1882, Jihlava, PSČ 587 33
<b>Osoby oprávněné za zadavatele jednat:</b>	Ing. Petr Pavlinec vedoucí odboru informatiky
<b>Kontaktní osoby:</b>	Ing. Petr Pavlinec
<b>Telefon:</b>	564 602 114
<b>E-mail:</b>	<a href="mailto:pavlinec.p@kr-vysocina.cz">pavlinec.p@kr-vysocina.cz</a>

<b>Zpracovatel</b>	
<b>Název:</b>	ICZ a.s.
<b>Adresa sídla:</b>	Hvězdova 1689/2a, 140 00 Praha 4
<b>Kontaktní osoby:</b>	Petr Tengler, Ing.Pavel Pačes
<b>Telefon:</b>	+420 244 100 111
<b>E-mail:</b>	<a href="mailto:petr.tengler@i.cz">petr.tengler@i.cz</a> , <a href="mailto:pavel.paces@i.cz">pavel.paces@i.cz</a>

# 1 Manažerské shrnutí

## 1.1 Předmět zakázky

Předmětem zakázky je dodávka koncepční studie, která popíše možné řešení problematiky digitalizace, ukládání a archivace el. dokumentů v regionálním měřítku a podmínkách kraje Vysočina.

Tato studie bude navazovat na projektový záměr technologické části tzv. eGON Center v kraji (technologické centrum kraje a ORP).

Cílem studie je její možné zobecnění na úroveň typového projektu regionální digitalizace archivace pro libovolný kraj ČR.

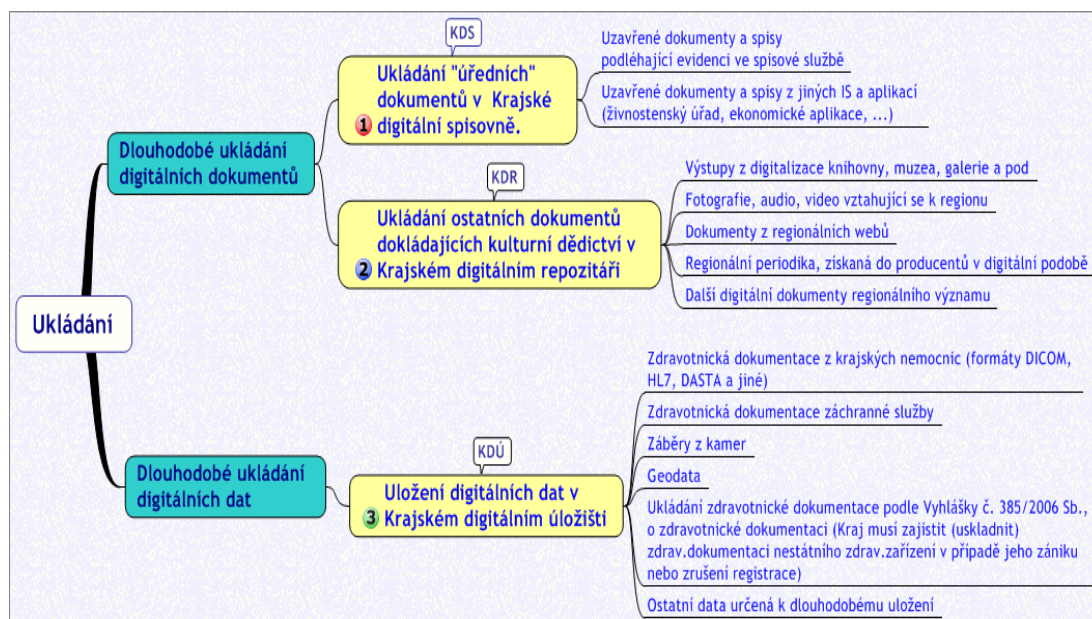
Z pohledu kraje jde pak o detailní analýzu zejména následujících procesů:

- zpracování a dlouhodobé uložení dokumentů orgánů veřejné správy
- záchrana, ochrana a zpřístupnění dokumentů z oblasti knižních fondů
- digitalizace a archivace cenných písemností a uměleckých děl
- zajištění dlouhodobé dostupnosti a čitelnosti el. záznamů
- archivace dalších dat

## 1.2 Nástin řešení

V agendách veřejné správy, a to na jednotlivých úrovních státní správy i samosprávy, vzniká již několik let velké množství elektronických dokumentů nejrůznějších typů uchovávaných na nejrůznějších typech nosičů. Některé z nich mají historický význam a potřebují zvláštní péči, aby se dochovaly jako svědectví minulosti. V současnosti se elektronické dokumenty zrovnoprávňují s papírovými a to ve všech fázích jejich životního cyklu. Je nutné podpořit celý životní cyklus elektronických dokumentů a dat. Kromě digitálních dokumentů je nutné bezpečně po neomezenou dobu zajistit také digitalizované dokumenty, vytvářené jednak pro umožnění snadnějšího přístupu badatelů k nim, ale také jako náhrady ohrožených nebo zanikajících fyzických archiválií.

Následující obrázek znázorňuje rozdělení problematiky ukládání :



### Terminologická poznámka:

Název této studie dle zadání zní: "Studie regionálního konceptu digitalizace a elektronické archivace kraje Vysočina".

Termíny "archivace" / "archivní" / "archivační" není na řadě míst v rámci této studie míněna archivace dle zákona 499/2004 o archivnictví a spisové službě, nýbrž čistě technicky pojaté řešení dlouhodobého důvěryhodného ukládání elektronických dokumentů či dat.

Hlavním důvodem používání této terminologie je celosvětový "de facto" model dlouhodobého ukládání digitálních dat OAIS (Open Archival Information System), který se do češtiny překládá jako "Otevřený

archivní informační systém", z něhož vycházíme při návrhu níže navrhovaných konkrétních subsystémů KDS, KDR a KDÚ pro Kraj Vysočina.

Dalším příkladem použití termínu "archivní" je základní technická komponenta systému OAIS, nazývaná v originále "Archival Storage", což se překládá do češtiny jako "Archivní úložiště" a realizuje se různými komerčními "archivními" systémy.

Dalším příkladem použití termínu "archiv" je t.zv. "Systém digitálního archivu", čímž míníme všechny komponenty systému OAIS mimo vlastní "Archivní úložiště"

Z pohledu služeb zajišťovaných Technologickým centrem kraje představuje ukládání a archivace tři větve:

1. Krajská digitální spisovna (KDS)
2. Krajský digitální repozitář (KDR)
3. Krajské digitální úložiště (KDU)

Z pohledu potřebného technického vybavení jsou první dvě služby implementovány jako dlouhodobý archiv podle modelu OAIS. Třetí služba je realizovaná jako technologicky zabezpečené úložiště souborů s definovaným přístupem pomocí běžných protokolů (ftp, nfs, cifs, http,...).

Tyto služby budou ukládat dokumenty a data následujících typů:

KDS - úřední dokumenty a spisy vzniklé jako produkt činnosti původců

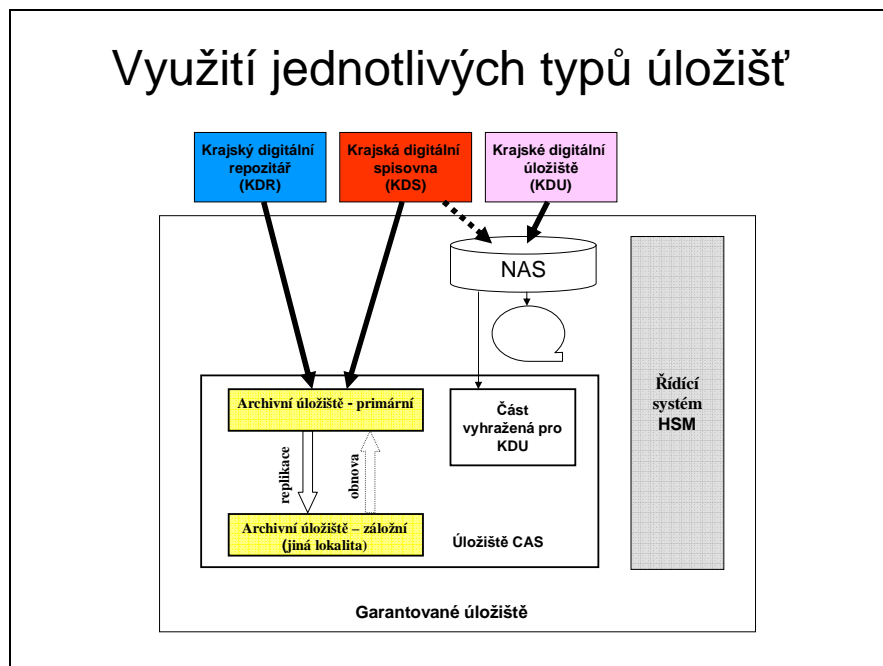
KDR - dokumenty převážně z oblasti kulturního dědictví regionu. To znamená vše, co lze považovat za dokumenty, má význam pro zachycení historie regionu a přitom dokumenty nevznikly jako produkt činnosti orgánů veřejné moci (knihy, filmy, fotografie, digitalizované umělecké předměty a sbírky)

KDU - jiná data a dokumenty, která pocházejí z činnosti informačních systémů kraje a je třeba je z nejrůznějších důvodů střednědobě až dlouhodobě ochránit proti ztrátě (zdravotní dokumentace, geodata, záznamy z kamer, údaje z provozu informačních systémů důležité pro jejich audit apod.)

KDS, KDR i KDÚ budou využívat jedno společné garantované úložiště (GÚ), realizované na technologii krajského Technologického centra. Garantované úložiště bude součástí Datového úložiště (Storage) typu Tier 3. Podle Studie proveditelnosti krajského Technologického centra bude tato úroveň úložiště realizována jako robustní NAS s HSM funkcionalitou.

Pro účely dlouhodobé archivace, popisované v této studii, bude tato technologie doplněna o ukládací technologii typu CAS (Content Adressable Storage), která bude rovněž začleněna pod řízení HSM. Nastavením ukládacích politik v rámci systému HSM tak bude možno řídit způsob uložení podle požadavků jednotlivých typů dat (a dokumentů) na diskové úložiště, pásku či na úložiště typu CAS, případně CAS/ WORM.

Využití jednotlivých typů úložišť ukazuje následující obrázek.



Krajská digitální spisovna i repozitář budou ukládat svoje objekty v podobě archivních informačních balíčků AIP podle modelu OAIS na úložiště typu CAS. Tato část CAS (tzv. Archivní úložiště) bude fyzicky replikována na dvě geografická místa (primární a záložní archivní úložiště) v rámci kraje.

Pro dokumenty uložené v KDS, patřící pod vybrané položky spisového plánu s velmi krátkou skartační lhůtou (např. jeden až dva roky) může být alternativně předepsáno ukládání do úložiště NAS, tak aby se do úložiště CAS ukládaly především dokumenty s dlouhou dobou životnosti. To znázorňuje čárkovaná šipka mezi KDS a NAS.

## 2 Rámec řešení

### 2.1 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.

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



**Krajské garantované úložiště** se skládá z Krajského Digitálního Úložiště (KDÚ) a Krajského Archivního Úložiště (KAÚ):

**Krajské 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.

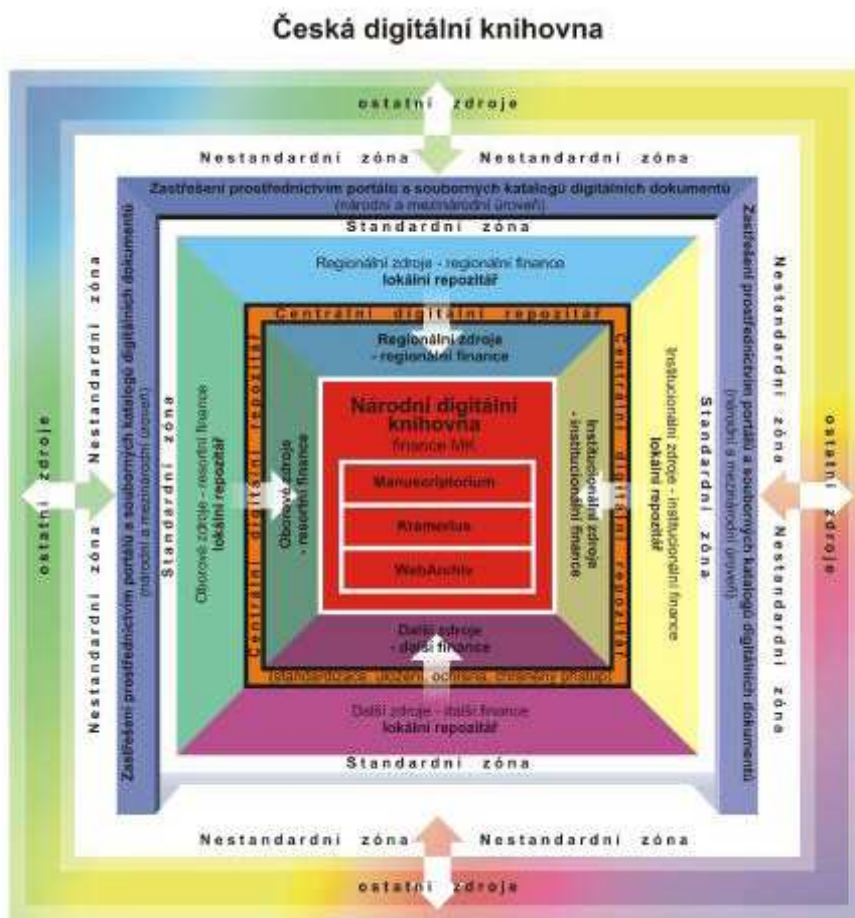
**K termínům "archivace" / "archivní" / "archivační" viz "terminologickou poznámku" v kapitole 1.2. "Nástin řešení"**

## 2.2 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 nejvzácnější 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 uchování (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.



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.

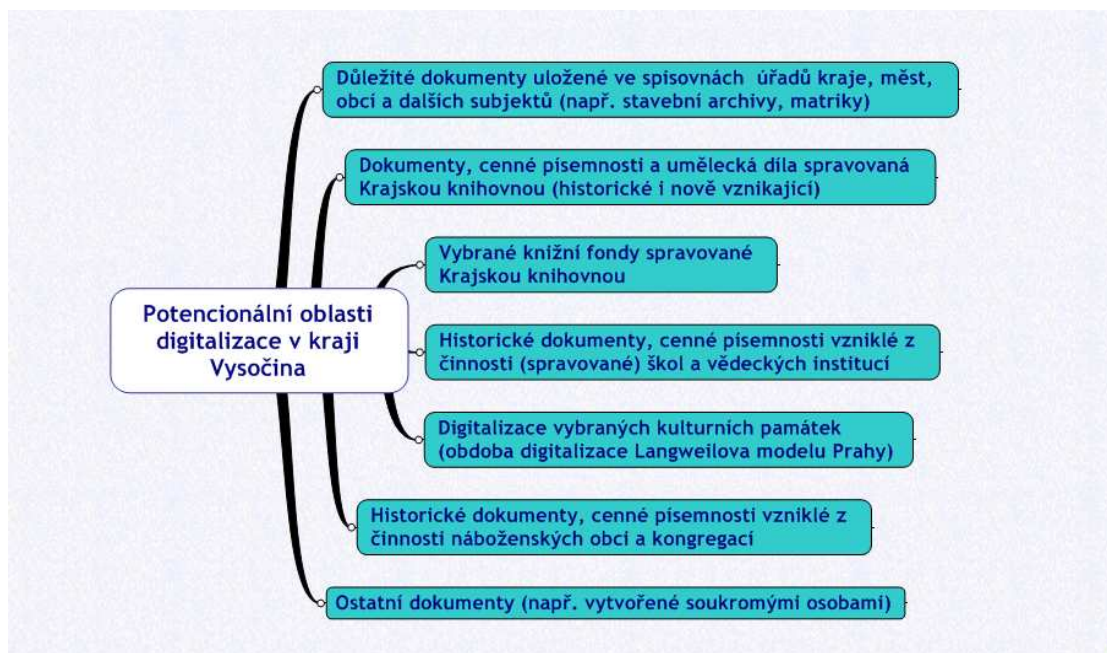
Vytvoření Národní digitální knihovny se zabývá digitalizací, uložením a zpřístupněním podkladů celonárodního významu v následujícím členění:

- 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.

Zatímco zdroje dokumentů pro projekt Národní digitální knihovny jsou zmapovány celkem podrobně, zdroje (podklady) digitalizace regionálního charakteru v rámci Krajských technologických center zatím nejsou přesně specifikovány. V obecné poloze by digitalizace měla být podle zadání zaměřena na zpracování dokumentů pro potřebu fungování úřadů a dále na záchranu, ochranu a zpřístupnění

dokumentů z oblasti knižních fondů, stavebních spisoven, zdravotnických spisoven nebo dokumentů významných svým obsahem či původem pro kulturní, politické, náboženské či jiné oblasti, kterým hrozí nebezpečí fyzického poškození či rozpadu v důsledku jejich častého používání.

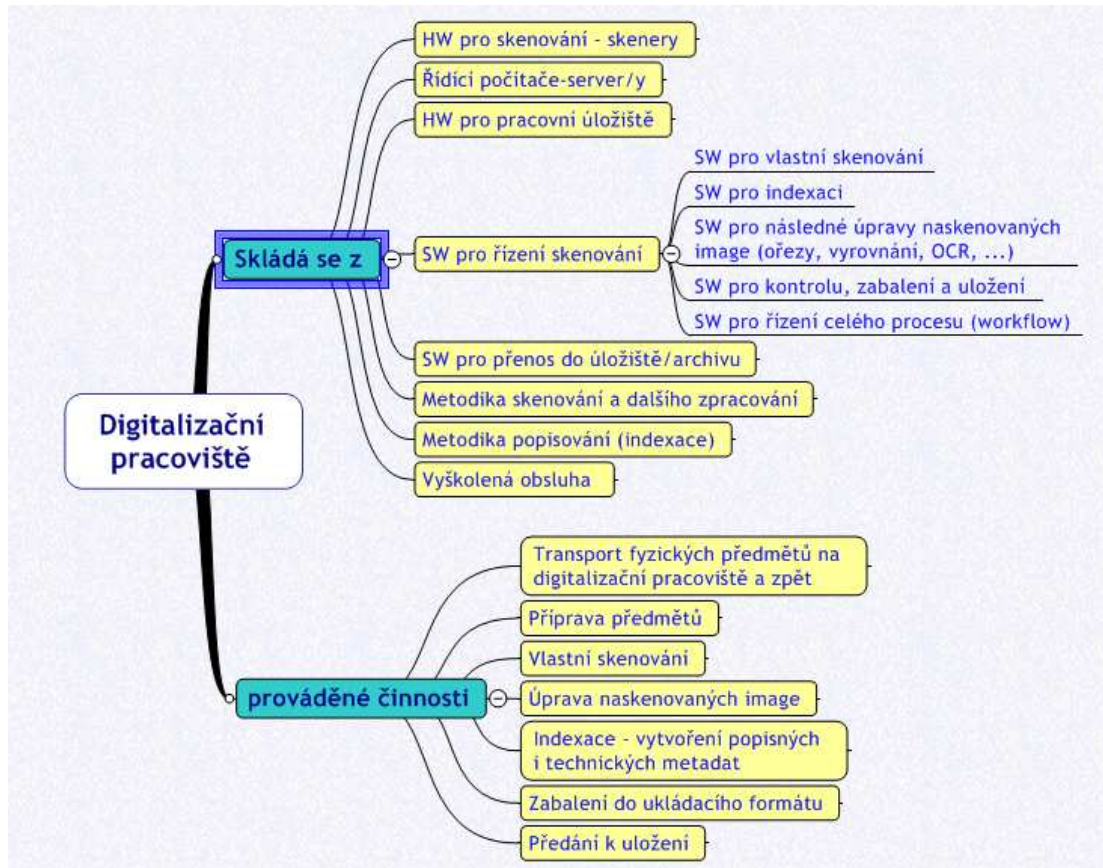


Konkrétní seznam zdrojů v rámci kraje Vysočina je mimo rozsah této studie. Digitalizace je pro účely této studie chápána jako jeden ze zdrojů digitálních dokumentů, který je nutno uchovávat. Vzhledem k celkovému značnému rozsahu a vysokým nákladům, bude vytvoření strategie postupu digitalizace regionálních zdrojů (nejen kraje Vysočina) zcela jistě podléhat nejen odbornému, ale především politickému rozhodnutí.

Pro zdárný průběh procesu digitalizace doporučujeme rozčlenit digitalizaci regionálních dokumentů na jednotlivé menší subprojekty podle typů zdrojů, které mají být digitalizovány. Cílem rozdělení na jednotlivé menší subprojekty je především zmenšení velikosti potřebných nákladů a tím i jednodušší alokace prostředků. Nezanedbatelným efektem rozdělení je pak i rychlejší sběr prezentovatelných výsledků a následné získání pozitivní zpětné vazby veřejnosti, odborné i politické reprezentace.

#### **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.



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
PM	PreservationMaster	Archivní soubor	JP2
MM	ModifiedMaster	Modifikovaný soubor	JP2
UC	UserCopy	Uživatelská kopie	PDF HT
OCR	TextFormat	Textový soubor	ALTO XML
DMD XML	DescriptiveMetadata	Popisná metadata	XML
AMD XML	AdministrativeMetadata	Administrativní metadata	XML

Z pohledu ukládání, resp. dlouhodobé archivace tyto soubory tvoří jeden digitální objekt („informační balíček“ podle modelu OAIS, viz kapitola 4), který má být uložen v (dlouhodobém) archivu.

Na úrovni kraje tedy bude existovat jedno, nebo několik digitalizačních pracovišť, která budou předávat výstupy své činnosti k archivaci buď do Národní digitální knihovny (protože digitalizují podklady celonárodního významu), nebo do Krajského digitálního repozitáře (protože se jedná o dokumenty regionálního významu). Všechny výstupy by měly být podchyceny v registru digitalizace, resp. v souborném katalogu. V případě potřeby bude možno digitalizační pracoviště sdílet pro oba projekty.



Proto doporučujeme pro projekty digitalizace na Krajské úrovni kompletně převzít metodiku digitalizačních pracovišť zpracovanou Národní knihovnou, včetně standardů pro výstupní datové formáty.

#### **Metodika digitalizace ostatních památek**

Problematikou digitalizace jiných než listinných objektů, například 3D skenování se zabývá řada pracovišť (muzea, galerie a vědecké instituce). Díky technologii nazývané 3D digitalizace se dá převést 3-rozměrný reálný objekt do digitální podoby. Zařízení, která nám tento převod umožní se nazývají 3D skenery. Na rozdíl od situace v digitalizaci listinných objektů zde existuje značná rozdílnost. Pokud z této činnosti vznikne jednotná metodika podobně jako pro digitalizaci listinných památek, doporučujeme ji využít i na regionální úrovni. Metodika by měla určit především formáty použitelné pro archivaci 3D dat, formáty metadatových popisů a typické postupy pro provedení digitalizace (především přípravu objektů, skenování, úpravu získaných dat, přípravu metadat, ...).

#### **Kompetence a odpovědnost**

Poněkud složitější situaci vidíme ve stanovení kompetencí a odpovědnosti za jednotlivé subprojekty digitalizace. Globální odpovědnost za projekt vybudování Technologického centra kraje je na krajské úrovni, potažmo na Ministerstvu vnitra. Jednotlivé zdroje dokumentů k digitalizaci však mohou spadat i pod jiné resorty. Proto doporučujeme při vzniku každého subprojektu digitalizace definovat nejen jeho předmět, způsob digitalizace, zdroje financování ale i přesně určit odpovědnost za výběr objektů a provedení digitalizace.

## 2.3 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 obrazová 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 uchováván). 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, navrhuje 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, navrhuje 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).

### 2.3.1 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 jednoduší 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 viz kapitola 4. 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 v kapitole 4.6.2.

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 zprocesovat. 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.

#### 2.3.1.1 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. (viz kapitola 4).

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 ze elektronické spisové služby (či jiného ERMS systému) a příjmu těchto dokumentů do repozitářů (spisoven, dlouhodobých úložišť), 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 (archivních) lhůt, které se pohybují v rozmezí 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.

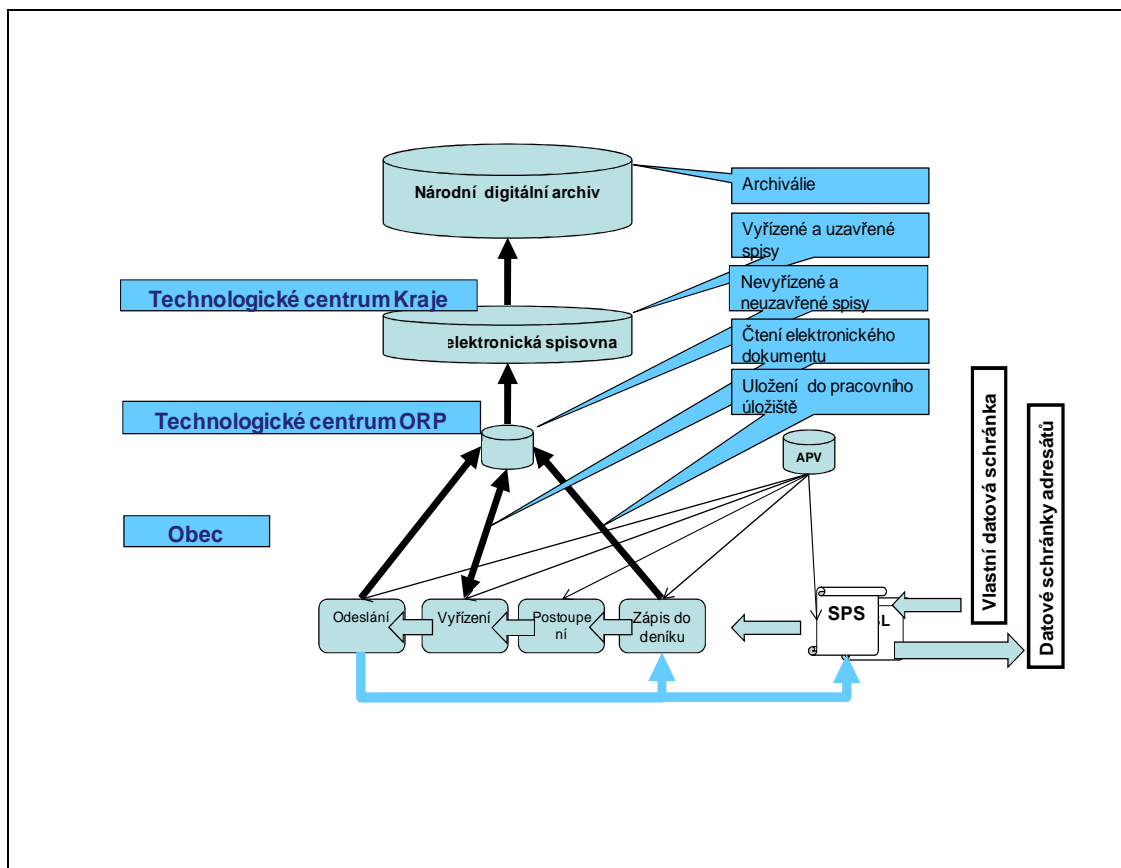
Úřední dokumenty a spisy vznikají a vyřizují se v různých IT systémech a aplikacích jako Podatelna, Elektronická spisová služba, Agendový systém pro stavební řízení, apod. Tyto systémy zajišťují příjem dokumentů, přípravu a vyřízení, odesílání a spojování do spisů v rámci správního řízení či jiných odborných 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. 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 (výběr je uveden v příloze studie) 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.





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](http://www.egoncentrum.cz/index.php?option=com_rubberdoc&view=doc&id=46&format=raw&Itemid=29).

### 2.3.1.2 Ad. 2. Archivace dokumentů kulturního dědictví regionálního významu.

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 kulturní památky, 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).

### 2.3.2 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 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.

Přítom 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 navrhujeme ukládat tato data v Krajském digitálním úložišti (KDU).

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.

## 3 Zdroje dat (Současný stav problematiky)

### 3.1 Krajská digitální spisovna

#### 3.1.1 Původci dokumentů

Současnými původci listinné Krajské spisovny jsou:

- orgány Kraje (hejtman, Rada, Zastupitelstvo, Krajský úřad včetně jeho organizačních jednotek). Zdrojem dokumentů jsou členové volených orgánů či zaměstnanci Krajského úřadu.
- zrušené příspěvkové organizace zřízené Krajem ( v současné době Školní statek Jihlava, Školní statek Havlíčkův Brod, Školní statek Telč, Dětská odborná léčebna Počátky, Školní statek Třebíč), mohou být další
- zrušené okresní úřady – dokumenty, které přešly v rámci spisové rozluky z okresních úřadů Jihlava, Pelhřimov, Havlíčkův Brod, Žďár n. Sáz. a Třebíč
- zrušená nestátní zdravotnická zařízení, převážně v případě úmrtí soukromého lékaře, jehož zdravotní dokumentaci jednotlivých pacientů má Kraj povinnost převzít

Budoucími původci KDS budou následující subjekty, u nichž vznikají digitální dokumenty:

- orgány Kraje (hejtman, Rada, Zastupitelstvo, Krajský úřad včetně jeho organizačních jednotek). Zdrojem dokumentů jsou členové volených orgánů či zaměstnanci krajského úřadu.
- Krajem zřízené příspěvkové organizace

- zrušená nestátní zdravotnická zařízení, převážně v případě úmrtí soukromého lékaře, jehož zdravotní dokumentaci jednotlivých pacientů má Kraj Vysočina povinnost převzít (týká se převážně převzetí elektronické zdravotnické dokumentace soukromých lékařů v případě jejich náhlého úmrtí)
- obce Kraje
- příspěvkové organizace zřízené obcemi Kraje

### 3.1.2 Zdrojové systémy

Do digitální spisovny Kraje se budou dostávat dokumenty převážně z elektronických systémů spisové služby (ESS) původců. Nebudou to ovšem jediné zdroje elektronických dokumentů. Dalšími zdroji jsou samostatné elektronické agendy (ERMS) produkující elektronické dokumenty různého typu podle metodik ministerstev. Dnes jsou to především ekonomické evidence (účetnictví), agendy životního prostředí, agendy živnostenských úřadů, agendy grantových agentur, v budoucnu možná i další<sup>1</sup>.

### 3.1.3 Odhad množství dokumentů

Při odhadu množství ukládaných dokumentů vycházíme z následujících metrik a předpokladů:

Metriky použité pro výpočet:	Elektronické dokumenty přenesené přes DS	Listinné dokumenty ke skenování	Důležité dokumenty skenované ve vysokém rozlišení	Multimediální dokumenty
Průměrný objem (velikost) jedné stránky A4 [kB]	25	80	50 000	
Průměrný objem (velikost) jednoho dokumentu [kB]	50	160	100 000	1 000 000

Z celkového počtu 1310 bude do KDS ukládat 100% původců.

Malých organizací je cca 3/4 z celkového počtu 605.

Přes datové schránky se přenesou cca 70 % z celkového počtu dokumentů.

Množství elektronických dokumentů doručených přes e-podatelný (e-mail) je zanedbatelné.

Z celkového počtu listinných dokumentů se budou skenovat asi 2/3.

V kalkulaci zanedbáváme objem metadat, protože je v poměru k obsahu velmi malý.

<sup>1</sup> Poznámka: Již nyní existuje zákonná povinnost vést matriční evidenci v ERMS, kde se uchovávají elektronické matriční knihy a současně vést matriky též ve formě listinné (viz zákon č. 301/2000 Sb., §6). Nelze vyloučit, že v budoucnu bude rozhodnuto, že matriční knihy se budou vést pouze elektronicky bez nutnosti duplicitní práce matrikářek. Od tohoto okamžiku bude nutno "digital born" matriční knihy dlouhodobě a důvěryhodně ukládat.

Na základě zkušeností z provozu spisových služeb u různých organizací odhadujeme následující počet a objem dokumentů zpracovávaných ročně:

Kategorie původce	Počet	Odhadovaný průměrný počet dokumentů za rok	Z toho elektronické dokumenty přenesené přes DS	Z toho dokumenty listinné ke skenování	Důležité dokumenty skenované ve vysokém rozlišení	Multimediální dokumenty	Celkový objem na jednu obci/org. [MB]	Procento původců, kteří budou ukládat do KDS	Objem za všechny původce v kategorii [MB]
Krajský úřad	1	300 000	210 000	60 000	2 000	20	240 100	100%	240 100
Malé obce (do 500 obyv.)	550	1 000	700	200	0	0	67	100%	36 850
Střední obce (500 - 1000 obyv.)	84	5 000	3 500	1 000	10	1	2 335	100%	196 140
Velké obce (s více než 1000 obyvateli včetně PO2 + PO3)	70	25 000	17 500	5 000	100	5	16 675	100%	1 167 250
Malé organizace (3/4 z celkového počtu 605)	454	1 000	700	200	5	0	567	100%	257 276
Velké organizace	151	5 000	3 500	1 000	10	0	1 335	100%	201 919
celkem [MB]									2 099 535
Celkem [TB]									2,1

Z odhadu vyplývá, že průměrný roční přírůstek digitálních dokumentů do KDS činí 2,1 TB dat.

### 3.1.4 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 podléhá schválení 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>.

### 3.1.5 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)

## 3.2 Krajský digitální repozitář

### 3.2.1 Původci dokumentů

Původci KDR budou libovolné subjekty v regionu Kraje, od nichž bude Kraj po dohodě s nimi vybírat a trvale uchovávat původně elektronické či digitalizované dokumenty kulturního dědictví regionálního významu. Kraj bude z těchto dokumentů vytvářet příslušné fondy či sbírky.

Bude se jednat například o tyto typy dokumentů:

- dokumenty, cenné písemnosti, umělecká díla a vybrané knihovní fondy spravované Krajskou knihovnou
- historické dokumenty a cenné písemnosti vzniklé z činnosti nebo spravované školami a vědeckými institucemi
- digitalizované sbírky muzeí a galerií
- 3D digitalizované vybrané kulturní památky (obdoba digitalizace Langweilova modelu Prahy)
- historické dokumenty a cenné písemnosti vzniklé z činnosti náboženských obcí a kongregací

- dokumenty vytvořené soukromými osobami
- webové stránky regionálního významu vytvořené libovolnými původci

### 3.2.2 Zdrojové systémy

Do KDR se dostávají elektronické dokumenty ve formě vstupních informačních balíčků (SIP), které mohou vytvářet z dodaných datových souborů (popisná metadata, obrázky) odborní pracovníci KDR nebo tyto balíčky vytvářejí podle dohody příslušní původci, kteří mají k dispozici uživatelské rozhraní KDR umožňující tvorbu těchto balíčků a zasílají je do KDR. V případě digitalizace na digitalizačním pracovišti budou balíčky SIP sestavovány v rámci exportu dat.

Protože jedním z nejvýznamnějších původců KDR bude Krajská knihovna, která bude digitalizovat kulturní dědictví regionálního významu z knihoven v regionu Kraje, bude mít mimo standardní možnosti vytvářet balíčky SIP i možnost vzdáleně přistupovat do KDR v rolích odborných pracovníků a ze zaslaných dokumentů vytvářet fondy či sbírky a dle potřeby upravovat popisná metadata.

Podobným zdrojem dokumentů bude i projekt digitalizace předmětů ze sbírek muzeí a galerií.

### 3.2.3 Odhad množství dokumentů

Metriky použité pro odhad:

Digitalizace knih – na jednom digitalizačním pracovišti se zpracuje průměrně 150 knih ročně, jedna kniha má v průměru 200 stran, archivní kopie má cca 16 MB na stránku výsledný objem 3 200 MB na jednu knihu. Odhadujeme, že do regionálního repozitáře se bude ukládat výstup z jednoho digitalizačního pracoviště.

Metriky použité pro výpočet:	Výstupy z digitalizace knih	Fotografie	Audio, video	Důležité dokumenty skenované ve vysokém rozlišení	Regionální periodika, získaná od producentů v digitální podobě	Běžné digitální dokumenty
Průměrný objem (velikost) jednoho dokumentu [MB]	3 200	5	1 000	100	4	2
Poznámky	1 kniha cca 200 stran, podle metodiky NK			Barevně ve vysoké kvalitě (600 DPI), uložené bez komprese	Např. Noviny Kraje Vysočina 3 - 4MB, Zpravodaj 1 - 2 MB, Zpravodaj oddělení mládeže a sportu do 500KB	25 stránek na dokument, 80 kB na stránku

V kalkulaci zanedbáváme objem metadat, protože je v poměru k obsahu velmi malý.

Kategorie	Počet	Počet za rok	Průměrný objem (velikost) jednoho dokumentu [MB]	Objem za rok
Výstupy z digitalizace - knihovny	Jedno digitalizační pracoviště	240	3 200	768 000
3D digitalizace - muzea a galerie		1	1 500 000	1 500 000
regionální tiskoviny-měsíčníky	5 ks	60	4	240
regionální tiskoviny-čtřmáctideníky	2 ks	48	4	192
regionální tiskoviny-týdeníky	1 ks	52	4	208
regionální tiskoviny-denníky	1 ks	240	4	960
regionální tiskoviny-občasníky		100	4	400
Fotografie		2 000	5	10 000
Audio, video		1 000	1 000	1 000 000
Důležité dokumenty skenované ve vysokém rozlišení		500	100	50 000
archivace vybraných webových stránek (změna cca 1x za týden)	20	1 040	4	4 160
Jiné digitální dokumenty		10 000	2	20 000
celkem [MB]				3 354 160
<b>Celkem [TB]</b>				<b>3,4</b>

Tato kalkulace je pouze příkladová, založená na laickém odhadu. Je třeba ji upravit podle konkrétních požadavků a podmínek.

Výsledkem odhadu je průměrný roční přírůstek digitálních dokumentů do KDR, který činí 3,4 TB dat.

### 3.2.4 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

### 3.2.5 Formáty elektronických dokumentů

Povolené formáty elektronických dokumentů pro uložení v KDR budou převážně stejné jako formáty pro KDS. Některé formáty, např. pro 3D skenování, bude třeba doplnit. 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, viz. kapitola 5.3.

## 3.3 Krajské digitální úložiště

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 nespecifikovaných dat.

Studie předpokládá uložení archivních dat v KDÚ ve formě souborů ukládaných původci prostřednictvím určených standardních souborových systémů nebo prostřednictvím aplikací umístěných v TC kraje (využití úložiště vrstvy Tier 3 dle „Studie proveditelnosti hostingového centra kraje Vysočina“). Úložiště neslouží pro přímé ukládání provozních dat původců na blokové úrovni.

Příkladem jsou soubory vzniklé záznamem z pouličních kamer, nebo 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.

### 3.3.1 Odhad množství dat

Metriky použité pro odhad:

#### Zdravotnická dokumentace PACS:

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

<b>Zařízení (Modality)</b>	<b>Typ dat</b>	<b>Odhad velikosti dat na jedno vyšetření</b>
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)

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.

Konzultacemi bylo zjištěno, že se objem dat ukládaných do PACS archivu pohybuje v průměru 1 - 1,5 TB na jednu nemocnici za rok. Tento údaj je třeba brát velmi orientačně, protože kromě velikosti se jednotlivé nemocnice velmi liší i vybavením a typem prováděných vyšetření. Navíc objem dat neustále stoupá, tak jak se zdravotnická zařízení postupně zdokonalují a roste jejich počet.

#### Zdravotnická dokumentace NIS, LIS a další IS:

Odhadujeme roční objem 200 GB na jednu nemocnici.

#### Záznamy z kamer:

Odhadujeme objem max 5 TB ročně.

#### Provozní logy:

Odhadujeme objem 1 MB za den.

#### Audio, video (např. záznamy z jednání zastupitelstva):

Odhadujeme objem 300 GB za rok.

Kategorie	Počet	Počet za rok	Průměrný objem (velikost) [MB]	Objem za rok [MB]
Zdravotnická dokumentace - PACS	5 nemocnic	5	1 200 000	6 000 000
Zdravotnická dokumentace - NIS, LIS, ...	5 nemocnic	5	200 000	1 000 000
Záznamy z kamer				5 000 000
Provozní logy	1 MB denně	365	1	365
Audio, video (např. záznamy z jednání zastupitelstva)				300 000
Ostatní	1 GB denně	365	1 000	365 000
celkem [MB]				12 665 365
<b>Celkem [TB]</b>				<b>12,7</b>

Výsledkem odhadu je průměrný roční přírůstek digitálních dat do KDU, který činí 12,7 TB dat.

### 3.3.2 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 (viz. kapitola 5.1.3).

### 3.3.3 Formáty dat

Pro KDU nelze předem stanovit žádná pravidla pro formáty dat.

## 3.4 Souhrnný odhad velikosti fyzického úložiště

Pro celkový objem fyzického úložiště jsme použili předchozí úvahy o ročním přírůstku jednotlivých úložišť a následující úvahy pro skartaci:

Pro Spisovnu KDS - po pěti letech se skartuje 80 procent dokumentů

Pro Repozitář KDR - skartace není žádná, nebo je zanedbatelná. Dokumenty zůstávají trvale.

Pro Úložiště KDU - Otázky skartace těchto dat zatím většinou zůstávají neřešeny. Pro výpočet jsme použili hodnotu po pěti letech se skartuje 30 procent dat.

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

Pro odhad nárůstu objemu byl použit lineární model i když víme, že je to velké zjednodušení. Objem dat se bude neustále měnit tak, jak bude postupovat akvizice původců, vývoj technologií, rozsah informatizace společnosti a pod.

Následující tabulka uvádí odhad objemu úložiště na 10 let.

	KDS - vstup [TB]	KDS - skartace [TB]	KDA - vstup [TB]	KDA - skartace [TB]	KDU - vstup [TB]	KDU - skartace [TB]	Objem archivního úložiště (pro KDS+KDA) [TB]	Celkový objem garantovaného úložiště [TB]
		80%		0%		30%		
1. rok	1,0				12,7		1,0	14
2. rok	2,1		3,4		12,7		6,5	32
3. rok	2,1		3,4		12,7		12,0	50
4. rok	2,1		3,4		12,7		17,5	68
5. rok	2,1	-1,7	3,4		12,7	-3,8	21,3	81
6. rok	2,1	-1,7	3,4		12,7	-3,8	25,1	94
7. rok	2,1	-1,7	3,4		12,7	-3,8	29,0	106
8. rok	2,1	-1,7	3,4		12,7	-3,8	32,8	119
9. rok	2,1	-1,7	3,4		12,7	-3,8	36,6	132
10. rok	2,1	-1,7	3,4		12,7	-3,8	40,4	145

Z příkladu výpočtu je vidět, že je třeba počítat s neustálým nárůstem objemu dat garantovaného úložiště a proto je třeba volit pro jeho realizaci zařízení s velkou škálovatelností a možností postupného rozšiřování<sup>1</sup>.

Pravidelné sledování přírůstku objemu dat a plánování velikosti úložiště patří k základním úlohám správce spisovny, repozitáře i úložiště.

Pro odhad kapacit elektronických spisoven a úložišť zatím nejsou žádné zkušenosti. Abychom se dobrali alespoň nějakých čísel, přijali jsme celou řadu zjednodušujících předpokladů, které jsme se snažili uvést. Na začátku projektu pro dlouhodobou archivaci je třeba udělat řádový odhad (jednotky, resp. desítky, resp. stovky TeraB. Nebo PetaB?) podle toho dimenzovat ukládací platformu s dostatečnou škálovatelností a počítat s tím, že bude třeba pravidelně doplňovat kapacitu médií (např.

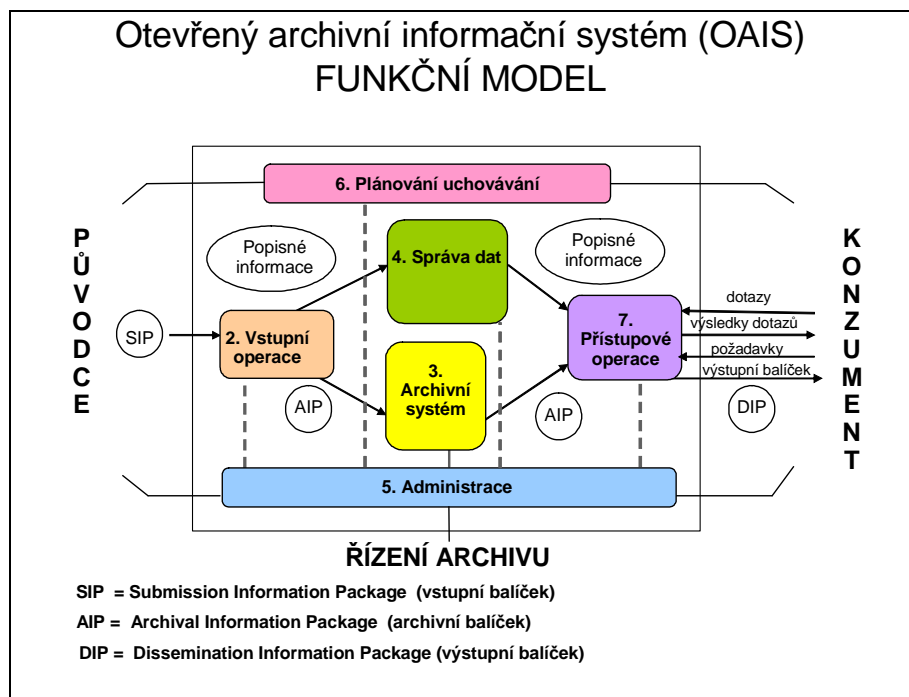
<sup>1</sup> Postupné rozšiřování je výhodné nejen z pohledu ceny (tak jak cena na uložení jednoho MB dat postupně klesá), ale také proto, že technologie úložišť se neustále vyvíjí. Je třeba počítat s tím, že jednou bude třeba vyměnit celou ukládací platformu, protože skončí její morální životnost.

disků) podle aktuálního zaplnění úložiště. Jednou za čas bude třeba doplnit další blok (skříň) platformy. Zda to bude blok stejné či jiné technologie bude třeba rozhodnout podle aktuální situace.

## 4 Požadavky na budoucí systém

### 4.1 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:



**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 uchovávání.

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 uchovávání"** (Preservation Planning). Tato služba odpovídá jak za plánování uchovávací 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 uchovávání" 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 technogickému 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

## 4.2 Rozdíly mezi jednotlivými subsystemy

Pro jednotlivé typy dokumentů a dat v subsystemech 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	<i>úřední dokumenty ve spisovně (KDS)</i>	<i>archivní dokumenty v repozitáři (KDR)</i>	<i>data v digitálním úložišti (KDÚ)</i>
<b>důvěryhodnost dokumentů či dat</b>	<p>Dokumenty musí být uloženy důvěryhodně - právní prokazatelnost. U dokumentů získaných z datových schránek je důvěryhodnost zaručena tímto institutem. U dokumentů, které tímto kanálem neprošly je vhodné je na vstupu do KDS opatřit elektronickým časovým razítkem.</p> <p>Upozornění: toto razítko pouze prokazuje existenci dokumentu a jeho metadat v daném čase a nemá nic společného s právní validitou ukládaného dokumentu.</p>	<p>Důvěryhodnost zajištěna postačujícím způsobem politikami GÚ v rámci TC, právní prokazatelnost nemá u archivních dokumentů význam.</p>	<p>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.</p>
<b>přístupová práva k dokumentům či datům</b>	<p>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 spisový plán.</p>	<p>Přístup musí být řízen systémem digitálního archivu, přístup musí být možno definovat podle nastavitelných politik na skupiny a role uživatelů.</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> některé dokumenty mohou být i obecně přístupné (veřejné)</li> <li><input type="checkbox"/> nebo přístup k nim je řízen <ul style="list-style-type: none"> <li>○ autorskými právy</li> <li>○ ochranou osobních údajů</li> </ul> </li> </ul>	<p>Přístup musí být řízen nastavením pravidel oprávnění technických prostředků úložiště (k jednotlivým adresářům filesystému)</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> přístup mají původci a jiné oprávněné osoby/orgány</li> <li><input type="checkbox"/> je nutno přihlížet k ochraně osobních údajů</li> </ul>

vlastnost	úřední dokumenty ve spisovně (KDS)	archivní dokumenty v repozitáři (KDR)	data v digitálním úložišti (KDÚ)
		<ul style="list-style-type: none"> <li>○ komerčními hledisky (zhodnocení nákladů na pořízení dokumentů)</li> </ul>	
<b>metadatové standardy</b>	<ul style="list-style-type: none"> <li><input type="checkbox"/> metadata stanovená Národním standardem pro elektronické systémy spisové služby (probíhá jeho schvalování)</li> <li><input type="checkbox"/> použít stejný formát jako pro ukládání do Národního digitálního archivu (probíhá jeho schvalování)</li> <li><input type="checkbox"/> povinná vazba na spisové plány původců</li> </ul>	<ul style="list-style-type: none"> <li><input type="checkbox"/> standard metadat stanovený Národní knihovnou pro knihovní systémy (probíhá jeho příprava a schvalování)</li> <li><input type="checkbox"/> možné vazby na číselníky stanovené Národní knihovnou (např. Celostátní databáze národních autorit vedená Národní knihovnou)</li> <li><input type="checkbox"/> základní archivní metadata používaná při budování archivních fondů a sbírek</li> <li><input type="checkbox"/> potencionálně může vzniknout požadavek i na využití standardů používaných v muzejnictví</li> </ul>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Metadatový standard nebude definován. Jednotlivá data-soubory nebudou opatřována metadaty. Globální metadatový popis bude dán strukturou katalogu KDÚ</li> </ul>
<b>způsob získávání metadat</b>	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.
<b>požadavky na vyhledávání a</b>	Specializované uživatelské rozhraní pro pracovníky původců, které musí	Specializované uživatelské rozhraní pro <ul style="list-style-type: none"> <li>○ administrátory repozitáře (kontrola</li> </ul>	Technické rozhraní pro správu úložiště, jednoúčelové rozhraní pro



<b>vlastnost</b>	<b>úřední dokumenty ve spisovně (KDS)</b>	<b>archivní dokumenty v repozitáři (KDR)</b>	<b>data v digitálním úložišti (KDÚ)</b>
<b>komfort pro konzumenty</b>	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"> <li>○ odborné pracovníky (kontrola vstupu dat, doplnění metadat a pod.)</li> <li>○ badatele</li> <li>○ širokou veřejnost. To musí být jednoduché a naprosto srozumitelné bez jakéhokoliv školení</li> </ul>	operátora KDÚ a další fyzické osoby s přidělenými přístupovými právy.
<b>existence vnitřních řídicích procesů</b>	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.
<b>spektrum technických formátů ukládaných dokumentů či dat</b>	Úzké, definováno standardy.	Úzké, definováno standardy a metodikou repozitáře.	Neomezené.
<b>způsob akvizice dokumentů či dat</b>	"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	"push" – sami původci ze své iniciativy zasílají data do KDÚ.

<b>vlastnost</b>	<b>úřední dokumenty ve spisovně (KDS)</b>	<b>archivní dokumenty v repozitáři (KDR)</b>	<b>data v digitálním úložišti (KDÚ)</b>
		digitalizace .	
<b>doba uložení</b>	Definovaná spisovým plánem max. cca 130 let.	Dlouhodobá – trvale.	Podle potřeby, dlouhodobost omezena životností formátů a SW.
<b>zdroje dokumentů či dat</b>	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.
<b>frekvence přístupů k dokumentům či datům</b>	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.
<b>počet současně pracujících konzumentů</b>	Relativně malý počet.	V některých časových úsecích vysoký zájem jak o vyhledávání, tak pro získávání dokumentů kulturního dědictví regionálního významu.	Malý počet.

## 4.3 Krajská digitální spisovna (KDS)

Navrhujeme vybudovat KDS podle modelu OAIS. Některé charakteristiky KDS v porovnání s ostatními subsystémy je možno vyčíst z tabulky uvedené v kapitole 4.2

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
- Nepřetržité zajištění základních charakteristik elektronických dokumentů uložených v KDS – životaschopnosti, čitelnosti, pochopitelnosti, autentičnosti a identifikace – podrobné vysvětlení viz kapitola 4.1
- 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 a současně i konzumenty KDS jsou subjekty vyjmenované v kapitole 3.1.1.

## 4.4 Krajský digitální repozitář (KDR)

Navrhujeme vybudovat KDR podle modelu OAIS. Některé charakteristiky KDR v porovnání s ostatními subsystémy je možno vyčíst z tabulky uvedené v kapitole 4.2

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 uchovávání a zpřístupňování elektronických dokumentů kulturního dědictví z regionu Kraje. Jedná se i o digitalizované dokumenty regionálního kulturního dědictví
- Nepřetržité zajištění základních charakteristik elektronických dokumentů uložených v KDS – životaschopnosti, čitelnosti, pochopitelnosti, autentičnosti a identifikace – viz kapitola 4.1
- 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:

- libovolné subjekty na území Kraje, u nichž vznikají digitální dokumenty regionálního kulturního dědictví
- významným původcem KDR je Krajská knihovna, která bude digitalizovat kulturní dědictví regionálního významu z knihoven a muzeí v regionu Kraje a jejich archivní kopie ukládat do KDR

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

- badatelé
- široká kulturní veřejnost

## 4.5 Krajské digitální úložiště (KDÚ)

Navrhujeme vybudovat KDÚ jako zabezpečený filesystém dostupný pomocí standardních protokolů a doplněný o katalog popisující uložená data podle jednotlivých vstupních kanálů.

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.

Některé charakteristiky KDÚ v porovnání s ostatními subsystemy je možno vyčíst z tabulky uvedené v kapitole 4.2

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

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 navrhuje pro takto uchovávaná data vytvořit katalog, kde bude udržován "globální" popis uložených dat, viz kapitola 5.1.3.

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.

## 4.6 Ostatní požadavky

### 4.6.1 Právní zakotvení KDR a KDÚ

Vzhledem k relativně nové problematice uchování 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 zřízení 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 „kopií“), 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 podíídit 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.

## 4.6.2 Proč je dobré použít OAIS model

Model OAIS představuje primární a odbornou veřejností akceptovaný standard pro dlouhodobé uchovávání 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 uchovávání 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 spisovny, 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 knihovníky, 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



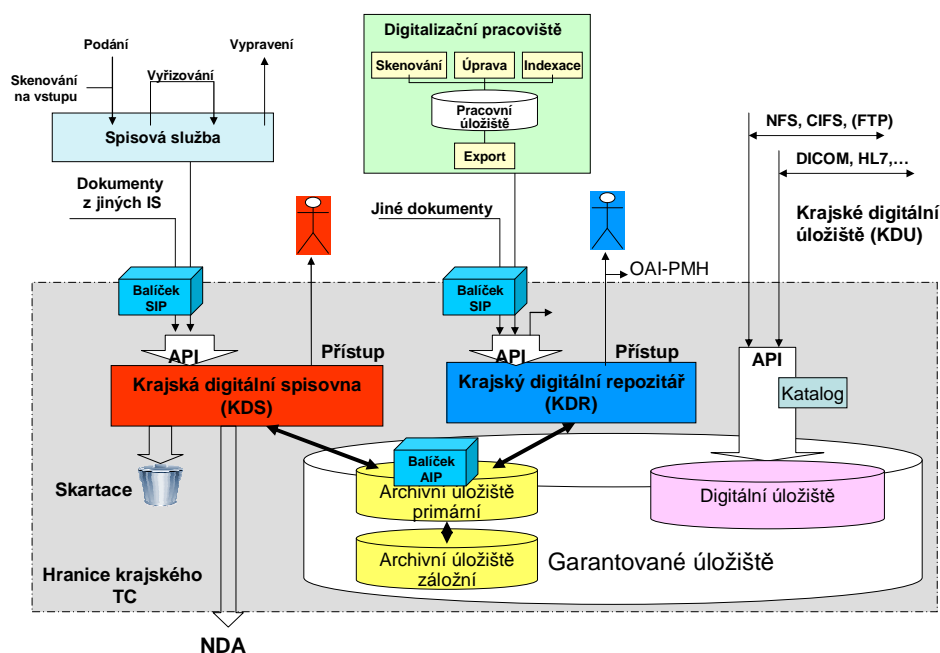
- Repozitář je považován za důvěryhodný, pokud lze prokázat jeho schopnost plnit určité funkce, a pokud tyto funkce splňují minimální dohodnutá kritéria, jež by měly splňovat všechny důvěryhodné repozitáře. Rozhodující je, že naplňování těchto kritérií musí být prokazatelné, což znamená, že dosažení 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.

## 5 Globální koncepce budoucího systému

### 5.1 Softwarová architektura

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

KDÚ je budováno jako systém pro zabezpečené uložení souborových dat bez evidence i kontroly metadat 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 využívající možností standardních souborových systémů.





### 5.1.1 Softwarová architektura KDS a KDR

Subsystémy 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.

System 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 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 implementováno primárně prostřednictvím technologie CAS (v případě vybraných klasifikací dokumentů v subsystému KDS i technologie 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.
  - 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

System 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

System eviduje veškeré přístupy k uloženým dokumentům a archivuje je.

## 5.1.2 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.
- 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 větší množině uživatelů nebo veřejnosti.
- Datové balíčky KDS budou zpřístupněny pouze omezené množině autorizovaných uživatelů původce.

### 5.1.3 Softwarová architektura KDÚ

Subsystem 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 subsystem nezajišťuje žádné další obslužné operace sloužící k zajištění konzistence, důvěryhodnosti a přístupnosti obsahu.

Na úrovni Technologického centra kraje může být v technicky a ekonomicky odůvodněných případech, 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 subsystemu 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 adresářová struktuře a uloženým datům jsou nastavena na základě definice oprávnění a kategorií podle údajů katalogu KDÚ.
- Předpokládá se realizace primárně s použitím technologie NAS s možností připojení CAS pro vybrané kategorie dat prostřednictvím HSM (viz. technologická architektura).

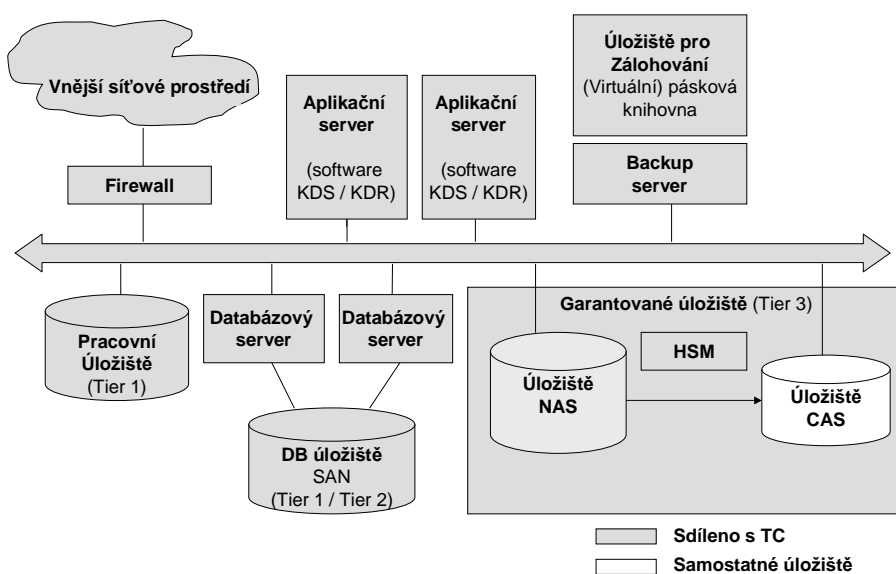
## 5.2 Technologická architektura

Systém by měl v maximální možné míře využívat hw a sw vybavení nasazené v rámci projektu Technologického centra kraje Vysočina (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 NAS a CAS (s použitím HSM) pro subsystémy KDS a KDR a KDÚ. Předpokládáme zajištění HA replikací 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



#### 5.2.1 Implementace subsystémů 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 TC K, přičemž bude možné využít i virtualizace těchto serverů.

K obslužným 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, jehož obsah bude kompletně replikován v záložní lokalitě. Obslužná aplikace pracuje s daty uloženými v CAS i po jejich uložení (čtení obsahu po vyžádání, procesy pro zajištění důvěryhodnosti a konzistence).

Pro vybrané dokumenty subsystému KDS (s krátkou skartační lhůtou - určené na základě konfigurace obslužného software KDS) je vhodné alternativně použít i úložiště NAS (Tier 3).

## 5.2.2 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 primárně úložiště NAS (Tier 3) současně s použitím technologie HSM. Pro vybrané kategorie uložených dat (například s požadavkem na uložení typu WORM) může být prostřednictvím HSM obsah přesunut do úložiště CAS, které může být takto využíváno současně s subsystémy KDS a KDR.

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 a CAS)
- pravidla pro přesun souborů mezi úrovněmi uložení prostřednictvím HSM v rámci NAS, případně uložení v CAS.

## 5.3 Uchovávací metoda (Digital preservation)

### 5.3.1 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 přečíst 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 uchování 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 rekonstrukcí 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.

### 5.3.2 Uchovávací metoda pro KDS a KDR:

Vzhledem k potenciální možnosti přebírání zkušeností, standardů i pokynů pro dlouhodobou archivaci zmíněnou v kapitole 2.3.1 navrhujeme 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 navrhujeme 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.



### 5.3.3 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, navrhujeme dvě opatření:

- omezení množství povolených formátů na vstupu,
- evidence použitých formátů a pravidelná kontrola ohrožení formátů, viz kapitola 4.5.

## 6 Závěry a doporučení

### 6.1 Dimenze konceptu budování a provozu KDS a KDR

Oba subsystemy, 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 předpověditelných 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.

## 6.2 Návrh etapizace a příslušných pilotních projektů

Vlastní realizaci KDS, KDR a KDÚ navrhujeme rozdělit na několik dílčích etap tak, aby docházelo k postupnému náběhu funkcionality. Důvodem rozdělení na etapy je jednak možnost průběžného financování projektu podle určených dodaných funkčních celků, ale především větší pravděpodobnost úspěšnosti realizace projektu a snadnější dosažení shody s požadavky a představami zadavatele. Též díky postupnému ověření dílčích částí je snadnější modifikovat řešení tak, aby bylo možné dosáhnout úspěšně funkčního řešení. Pilotní provoz systémů KDS, KDR a KDÚ vyžaduje instalaci základního HW a SW Technologického centra kraje včetně Garantovaného úložiště alespoň v minimální konfiguraci.

Před každou etapou je nutno podle aktuálního stavu znalostí současného stavu a zamýšlenými cíli aktualizovat plány podle metodiky PLATTER (viz kapitola 6.1.). Tuto metodiku je vhodné použít v nezávislých stromech pro KDS i KDR a přiměřeně i pro KDÚ.

### 6.2.1 První etapa

Tato etapa bude zahrnovat pilotní řešení KDS pro dokumenty od jednoho původce, nejlépe přímo Krajského úřadu. V rámci etapy bude dodáno a implementováno i úložiště typu CAS - jeho primární část.

Bude testován příjem dokumentů, jejich uložení a zpřístupnění. Budou otestovány základní principy zabezpečení dokumentů, import a aktualizace spisového plánu tohoto původce.

Následně bude otestováno připojení několika vybraných původců přebírání dokumentů včetně importu a aktualizace spisových plánů jednotlivých původců.

Po vyhodnocení pilotního běhu vznikne finální plán pro náběh do rutinního provozu KDS a připojení dalších původců.

### 6.2.2 Druhá etapa

V této etapě bude realizován systém KDÚ, který pracuje nezávisle na KDS a proto se může druhá etapa s první časově překrývat. I v této etapě doporučujeme připravit nejprve pilotní provoz KDU s jedním vhodným typem dat, na němž bude otestováno ukládání a výběr dat a práce s katalogem.

### 6.2.3 Třetí etapa

Cílem této etapy je realizace KDR. SW KDR využívá identických nebo modifikovaných modulů jako KDS a proto je možno s výhodou využívat zkušeností se zaváděním KDS.

U KDR je nutné v pilotním provozu otestovat především práci s metadatovými knihovnami a archivními standardy, které se liší od popisných metadat používaných u KDS, kde těžiště leží především ve spisových plánech.

Z tohoto důvodu doporučujeme zahájit pilotní projekt KDR v součinnosti s Krajskou knihovnou jako původcem, který má bohaté zkušenosti s pořizováním metadat k digitalizovaným knihovním fondům.

### 6.2.4 Čtvrtá etapa

Cílem etapy je dokončení přechodu celého řešení do rutinního provozu. Bude třeba zajistit veškeré vlastnosti deklarované v modelu OAIS. Tím je v první řadě zprovoznění záložního úložiště CAS v jiné lokalitě než je umístěno primární úložiště.

Dále je třeba velmi podrobně dopracovat a konkretizovat veškeré pracovní postupy a metodiky viz kapitola 6.1. Na závěr této etapy doporučujeme provést audit celého řešení archivace.

## 6.2.5 Rutinní provoz

Vlastní provoz digitální spisovny a repozitáře doporučujeme rozdělit na tzv. provozní cykly. Provozní cyklus je časově omezen implementací digitálního archivu na jednom typu technologie (HW+SW). Provozní cyklus končí (a začíná) migrací celého řešení na novou technologii.

Podpora existujícího technologického prostředí, ve kterém je digitální archiv provozován, již končí a je nutné celý archiv na převést novou technologii. Jen velmi těžko lze predikovat dopředu vývoj jakýchkoli technologií, počítačové nevyjímaje. Pro rozumné plánování je vhodné předpokládat, že provozní cykly budou trvat 10 - 15 let. Je však možné, že jeden cyklus může trvat i déle. Mezi dvěma cykly je z pohledu provozu nutno provést přípravu nového řešení, jeho odzkoušení, přípravu migrace dat a její odzkoušení. Převod/migrace dat při které se převedou data ze starého systému na nový. Je tedy nezbytné počítat s tím, že určitou dobu budou vedle sebe existovat oba systémy.

Tak jak budou dokumenty postupně přibývat, poroste kapacita potřebných úložišť. V každém provozním cyklu se tedy budou vyskytovat okamžiky, kdy bude třeba doplnit úložiště o další jednotky. Toto postupné doplňování úložné kapacity je opět navrhováno především z důvodů ušetření nákladů. Kromě toho že průběžně klesají ceny zařízení, stoupá jejich použitelná kapacita takže je finančně nevýhodné nakupovat příliš velkou kapacitu dopředu. Časový okamžik potřeby doplnění úložné kapacity lze opět velmi těžko dopředu predikovat (záleží na počtu dokumentů vstupujících do archivu a na jejich velikosti. Z pohledu plánování finančních prostředků je vhodné uvažovat s doplněním každých pět let. Zároveň s doplněním úložných kapacit je třeba počítat s přechodem na nové verze použitého základního software a s případným upgrade HW. Jako příklad lze uvést postupný vývoj verzí operačních systémů či databází, který většinou nevyžaduje kompletní migraci řešení nebo postupný vývoj výkonnějšího HW,

Plánováním provozních cyklů, jakož i plánováním doby doplnění ukládací kapacity se musí zabývat určenými pracovníky IT. Tyto činnosti je třeba zajišťovat po celou dobu života digitálního archivu.

## 6.3 Rizika projektu

Protože se jedná o zcela nový předmět činnosti na úrovni Kraje (pokud pomineme vedení listinné spisovny), je nutné předem vydefinovat základní aktivity, které je nutné v rámci implementace projektu uskutečnit a k nim si stanovit potenciálně vyplývající hrozby a jejich důležitost.

Posláním této kapitoly není zpracování risk analýzy komplexním způsobem, ale pouze zdůraznit klíčové faktory úspěšnosti zavedení a provozování digitálního úložiště, na které je nutno pamatovat.

Součástí strategie musí být i vypracování plánu resp. metodiky, která zajistí neustálou aktualizaci analýzy rizik.

Pro bezproblémový provoz je nezbytné formalizovat bezpečnostní pravidla ať pro fyzické zajištění tak i pro ICT a personál. Pro tuto oblast doporučujeme použít normu ČSN ISO/IEC 27001. Na problematiku vlastní digitální archivace doporučujeme provést interní analýzu rizik pomocí nástroje DRAMBORA.

### Hlavní rizika a návrhy opatření

Rizika jsou rozdělena do dvou skupin podle stavu budování digitálního úložiště:

- před provozním stavem (základní strategická rozhodnutí a „operativní“ rozhodnutí při řešení přípravy budování Krajského digitálního úložiště),
- v provozním stavu.

Při četbě následujících tabulek doporučujeme srovnání s kapitolou 6.1.

## Rizika před provozním stavem

Základní rozhodnutí (strategická rozhodnutí)	
Popis rizika	Návrh opatření
Nesoulad sdílení stejné vize mezi orgány kraje, vedení krajského úřadu a zaměstnanci	Sjednotit strategickou vizi Uskutečnit informativní schůzky Zajistit distribuci informací
Nebudou zajištěny dostatečné finančních zdroje a s tím související udržitelnost projektu delší než je minimální doba uvedená v IOPu (tzn. 5 let po ukončení realizace projektu).	Využití čerpání finančních prostředků z IOP. Naplánování prostředků na provoz digitálního archivu na dlouhou dobu. Povinné vytváření rozpočtových rezerv
Nejistý právní statut Krajského digitálního repozitáře a úložiště	Oponování připravovaného zákona, ukládání kopií digitálních archiválií předávaných NDA. Smluvní zakotvení vztahu s původci a badateli. Zhotovení právního auditu závazkových vztahů
Zvolení nestandardního modelu ukládání a správy dokumentů	Rozbor s pracovníky Národního archivu resp. Národní knihovny Použití celosvětově i lokálně srozumitelného modelu OAIS
Nedostatečné zajištění kvalitních lidských zdrojů pro jednotlivé etapy koncepce, vývoje, produkce a provozu.	Školení uživatelů, motivace k působení na projektu Zajistit si metodickou pomoc od NDA, NDK
Rizika při řešení budování Krajského digitálního úložiště (operativní rozhodnutí)	
Nekvalitní zpracování zadávací dokumentace	Postupovat v souladu se zpracovanými studiemi Sestavení konzultační a oponentní skupiny
Nesprávná organizace veřejné zakázky a z toho vyplývající zdržení, neadekvátní protiplnění apod.	Pečlivá příprava zadávací dokumentace Nastavení jasných výběrových a kvalifikačních kritérií např. bez možnosti dvojznačného výkladu
Nedostatečné zajištění bezpečnosti informací	Určit standardy a zodpovědnou osobu Začlenit problematiku archivace a ukládání do bezpečnostního projektu pro TC. Prověřování dostupných informací a reportů za účasti širšího sboru odpovědných pracovníků Proaktivní přístup
Podcenění zajištění legislativních požadavků	Určení odpovědných a poučených osob za sledování této problematiky Prověřování procesů a nastavení povinného reportování Zvažovat i pravděpodobné budoucí scénáře (např. prohlášení vlivných pracovníků NA či MV)

Základní rozhodnutí (strategická rozhodnutí)	
Popis rizika	Návrh opatření
Technické problémy mj. při řešení rozhraní na systémy původců archiválií a další systémy Stabilita nových standardů NDA, NDK	Využití rozhraní pro předávání archiválií definované pro NDA. Zajistit si metodickou pomoc od NDA, NDK V rámci kvalifikačních kritérií požadovat doložit technickou znalost nebo garantovanou podporu ze strany dodavatele systému, který bude připojován

### Rizika v provozním stavu

Popis rizika/aktivity	Návrh opatření
Zanedbání aktualizace systému jak po technické, metodické nebo legislativní stránce	Zajistit si metodickou pomoc od NDA, NDK Zajištění supportu a maintenance od dodavatelů jednotlivých součástí Účastnit se odborných akcí (školení, kongresů apod.) Být alespoň pasivně zapojen do členství v odborných organizacích Sledování připravované legislativy a plánování rozvoje
Nedostatečné provádění běžné správy a údržby, neúplné vedení dokumentace	Stanovit požadované postupy, metodiky, standardy a odpovědné osoby Pravidelná kontrola dodržování stanovených postupů a metodik
Nezajištění dostatečných finančních zdrojů	Vždy v rozpočtu uplatnit požadavky na dostatečné finanční zdroje Zpracování do každoročně sestavovaných rozpočtových výhledů a dlouhodobých plánů
Nesprávné řízení lidských zdrojů	Definování personálních standardů Trvání na účasti na pravidelných školení a zvyšování si kvalifikace Zajistit přiměřenou zastupitelnost na všech funkčních místech Zvážit využití externích zdrojů

## 7 Přílohy

### 7.1 Spisový plán Krajského úřadu Vysočina – vybrané položky

Do následující tabulky byly vybrány ty položky spisového plánu, jejichž skartační lhůta je delší než 10 let, případně některé zajímavé z těch, u nichž skartační lhůta začíná plynout nikoliv ihned po uzavření dokumentu či spisu, ale až po určité době, která je stanovena poznámkami ve spisovém plánu – jako t.zv. "spouštěcí události" – viz dále. Díky tomuto faktoru se skutečná skartační lhůta může prodloužit někdy až o další desítky let nebo i natrvalo. I z tohoto důvodu je důležité zajistit trvalé důvěryhodné uložení v digitální spisovně, které nemůže být standardními ERMS systémy zajištěno.

Do následujícího přehledu byly zařazeny typy dokumentů, u nichž se díky předpokládanému rozvoji použití datových schránek změní jejich povaha z dokumentů listinných na dokumenty digitální – "digital born".

Zatím nejdelší skartační lhůtu mají matriční knihy narození A/100 (přičemž skartační lhůta začíná u nich běžet po uzavření svazku) a Zdravotnická dokumentace převzatá od zdravotnických zařízení zrušených bez právních nástupců S/100. V současnosti vedou praktičtí lékaři či specialisté zdravotní dokumentaci velmi často v elektronické formě a teprve při předávání pacientovi nebo z důvodů archivačních ji tisknou na papír. Při náhlém úmrtí lékaře existuje tedy pouze "digital born" forma této dokumentace. Neexistuje-li nástupce tohoto lékaře, je nutno převzít tuto elektronickou dokumentaci do Krajské spisovny a pečovat tam o ní 100 let.

**Spisový plán Vysočina, vybrané potencionální elektronické dokumenty se skartačními lhůtami delšími než 10 let a v některých případech navíc též se spouštěcí událostí.**

<b><i>položky spisového plánu:</i></b>	
56 Smlouvy (nejsou-li nedílnou součástí příslušných dokumentů a spisů):	
56.1 obchodní	V/5 <sup>1)</sup>
56.2 nájemní	V/5 <sup>1)</sup>
56.3 majetkoprávní	A/5 <sup>1)</sup>
56.4 kolektivní	A/5 <sup>1)</sup>
56.5 o spolupráci	A/5 <sup>1)</sup>
56.6 veřejnoprávní	A/5 <sup>1)</sup>
56.7 darovací	V/5 <sup>1)</sup>
56.8 o dílo	S/5 <sup>1)</sup>
56.9 ostatní	V/5 <sup>1)</sup>
118.1.2.1 dokumenty s údaji pro účely důchodového pojištění	S/50
119.3 Výběrová řízení na obsazení pracovních míst	S/30
121.8 Mzdové listy	S/50
201.1 Vynětí zemědělské půdy ze zemědělského půdního fondu	A/15

209.1 Rozhodnutí o žádosti o vydání integrovaného povolení	V/50 <sup>1)</sup>
209.2 Uložení nápravných opatření	V/50 <sup>1)</sup>
209.3 Rozhodnutí o zastavení provozu zařízení nebo jeho částí	V/50 <sup>1)</sup>
221.2 Lesní hospodářské plány	A/20 <sup>1)</sup>
221.7 Rozhodnutí o odnětí lesních pozemků nebo omezení jejich využívání pro funkci lesa	A/20
231.2 Vodoprávní povolení, souhlasy a jiná opatření vodoprávního úřadu ve správním řízení	A/20
231.8 Vodoprávní evidence	A/50
231.10 Hraniční vody	A/20
231.12 Revitalizace vodních toků	A/20
233.3 Plán rozvoje vodovodů a kanalizací	A/20
233.4 Majetková evidence vodovodů a kanalizací	A/50
277.1.4 Osvědčení o zkouškách odborné způsobilosti v nákladní dopravě	S/50
277.2.4 Osvědčení o zkouškách odborné způsobilosti k dopravě osob	S/50
452.5 Výkonové výkazy škol a školských zařízení	V/20
454.9 Nostrifikace a uznání rovnocennosti vysvědčení vydaných zahraničními školami	S/50
526.2 Registrace nestátních zdravotnických zařízení	V/50
534 Zdravotnická dokumentace převzatá od zdravotnických zařízení zrušených bez právních nástupců	S/100
559.3 Odvolání proti rozhodnutí na úseku dávek pomoci v hmotné nouzi	S/15
559.4 Odvolání proti rozhodnutí o příspěvku na péči	S/15
606.1 Nabytí státního občanství udělením	A/50
606.2 Nabytí státního občanství jiným způsobem	A/50
606.3 Pozbytí státního občanství	A/50
606.4 Osvědčování a potvrzování státního občanství	A/50
607.1 Matriky (druhopisy) a sbírky listin:	
607.1.1 narození	A/100 <sup>6)</sup>
607.1.2 manželství	A/75 <sup>6)</sup>
607.1.3 úmrtí	A/75 <sup>6)</sup>
608 Odvolání proti rozhodnutí matričního úřadu o změně jména a příjmení	A/75
701 Odvolání na úseku státní sociální podpory	S/15
702 Rozhodování na úseku státní sociální podpory	S/15
706 Prominutí podmínky trvalého pobytu pro účely státní sociální podpory	S/15

**Vysvětlivky:**

1) Skartační lhůta začíná plynout po ztrátě platnosti dokumentu

6) Skartační lhůta začíná plynout po uzavření svazku.

Dle Věstníku MV č. 76/2009 kterým se zveřejňuje národní standard pro elektronické systémy spisové služby se "**spouštěcí událostí**" rozumí kritéria, kterými je stanoven začátek plynutí skartační lhůty. Spouštěcí událost je – pokud jí není uzavření dokumentu nebo skupiny dokumentů – vyjádřena ve spisových a skartačních plánech, a to obvykle poznámkou.



**Další důležité a často používané spouštěcí události skartační lhůty:**

- 1) Skartační lhůta začíná plynout po ztrátě platnosti dokumentu.
- 2) Skartační lhůta začíná plynout po vyhodnocení dokumentu.
- 3) Skartační lhůta začíná běžet po roce, v němž byla výpočetní a jiná technika naposledy použita, nebo po ukončení používání příslušného software.
- 4) Skartační lhůta u dokumentů celostátně vyhlášeného referenda začíná plynout po vyhlášení výsledků referenda prezidentem republiky ve Sbírce zákonů, popřípadě po vyhlášení nálezu Ústavního soudu, kterým rozhodl, že postup při provádění referenda nebyl v souladu s ústavním zákonem o referendu o přistoupení České republiky k Evropské unii nebo zákonem vydaným k jeho provedení s povinností zachování tří nepoužitých hlasovacích lístků pro referendum pro uložení v příslušném archivu. Skartační lhůta u dokumentů krajského referenda začíná plynout po vyhlášení výsledků referenda s povinností zachování tří nepoužitých hlasovacích lístků pro referendum pro uložení v příslušném archivu.
- 5) Skartační lhůta začíná plynout po uzavření knihy, u automatizované evidence po uzavření roku.
- 6) Skartační lhůta začíná plynout po uzavření svazku.
- 7) Skartační lhůta začíná plynout po vyhlášení výsledků voleb.
- 8) Dokumenty se zařadí do skartačního řízení poté, kdy Ministerstvo financí vydá samo nebo na žádost účetní jednotky souhlas k jejich vyřazení.
- 9) Skartační lhůta začíná plynout po zrušení zařízení.
- 10) Skartační lhůta začíná plynout po pořízení nové evidence.
- 11) Skartační lhůta začíná plynout po zrušení platnosti starých.
- 12) Skartační lhůta začíná plynout po roce, v němž byla technika vyřazena.
- 13) Skartační lhůta začíná plynout po uzavření smlouvy, její změny, nebo od zrušení zadávacího řízení.
- 14) Agenda se uchovává a předává jako celek za volební období.
- 15) Skartační lhůta začíná plynout po ukončení právní ochrany

## 7.2 ZKRATKY A POJMY

### 7.2.1 Zkratky

Zkratka	Význam	Ekvivalent / Popis
AIP	Archive Information Package	Archivní informační balík, zahrnující ukládaný obsah a odpovídající popisné informace pro uchovávání (archivní a technické informace), ukládané uvnitř OAIS.
CAS	Content addressable storage	Paměťové úložiště určené k dlouhodobému ukládání neměnného obsahu. Generuje globálně jedinečné jméno pro každý uložený objekt, které je nezávislé na běžné výpočetní infrastruktuře, organizaci, umístění nebo



Zkratka	Význam	Ekvivalent / Popis
		technologii. Objekt je vždy přístupný pod tímto jménem. Důsledkem toho je, že objekty se mohou v průběhu času libovolně přemísťovat, měníce své fyzické nebo technologické umístění, aniž by to mělo vliv na jejich přístupnost.
CIFS	Common Internet File System	Protokol pro sdílení souborů a tiskáren
DASTA	Datový standard definovaný Ministerstvem zdravotnictví	Slouží k předávání dat mezi zdravotnickými informačními systémy, je využíván v každodenní praxi již více jak deset let a je zabudován do všech současných významných (českých) zdravotnických informačních systémů.
DICOM	Digital Imaging and Communications in Medicine	Standard pro manipulaci, ukládání, tisk a přenos informací uložených v lékařských obrazových materiálech. Zahrnuje definici formátů souboru a síťový komunikační protokol.
DIP	Dissemination Information Package	Informační balík odvozený z jednoho nebo více AIPů, posílaný uživatelům (badatelům) jako odpověď na žádost o poskytnutí informace z OAIS.
ERMS	Electronic records management system	Elektronický systém spisové služby je informační systém určený ke správě dokumentů ve smyslu ustanovení § 2 písm. k) zákona, s použitím jeho § 63 odst. 3.
ESS	Elektronická spisová služba	Viz pojmy – elektronická spisová služba
FTP	File transfer protokol	Internetový protokol určený pro přenos souborů mezi počítači, na kterých mohou být provozovány rozdílné operační systémy.
GÚ	Garantované úložiště	
HL7	Health Level Seven	Nezisková organizace zabývající se vývojem standardů v oboru zdravotnické péče. Zde framework a s ním související standard pro výměnu, integraci, sdílení a získávání zdravotnických informací v elektronické podobě.
HSM	Hierarchical Storage Management	HSM je technika ukládání dat, která automaticky přenáší data mezi vysoko a nízko nákladovými ukládacími médii. Příkladem HSM scénáře může být např., že datové soubory, které jsou často používány, zůstávají

Zkratka	Význam	Ekvivalent / Popis
		uchovávané na discích, ale pokud nejsou užívány po delší časové období (typicky několik měsíců), mohou být migrovány na pásky. Jestliže jsou tato data umístěna na páskách a uživatel je potřebuje znovu použít, data jsou automaticky přenesena zpátky na disky.
HTTP	Hypertext transfer protokol	Internetový protokol určený původně pro výměnu hypertextových dokumentů ve formátu HTML.
HW	Hardware	
IOP	Integrovaný operační program	Integrovaný operační program je určen pro řešení problémů národního i regionálního charakteru za finanční podpory Evropské unie. Cílem IOP je např. podpora modernizace a zefektivnění činností a procesů v oblasti veřejné správy, veřejných služeb a územního rozvoje - rozvoj informačních technologií, digitalizace vybraných datových zdrojů, eGovernment, zlepšování infrastruktury v oblasti sociálních služeb, veřejného zdraví, služeb zaměstnanosti a služeb v oblasti bezpečnosti..... Typická míra spolufinancování z fondů EU je stanovena maximálně na 85 % z celkových veřejných výdajů.
KDR	Krajský digitální repozitář	
KDS	Krajská digitální spisovna	
KDÚ	Krajské digitální úložiště	
LIS	Laboratorní informační systém	
NAS	Networked Attached Storage	Zařízení pro uložení dat a centrální sdílení dat mezi různými platformami. Skládá se ze dvou základních částí: bezpečného datového úložiště a zařízení (NAS hlava, nebo NAS box) se spec. operačním systémem, optimalizovaným pro rychlou a spolehlivou práci se soubory. Klientovi se celé NAS zařízení jeví jako jeden fyzický disk.
NDA	Národní digitální archiv	
NDK	Národní digitální knihovna	
NFS	Network file system	Internetový protokol pro vzdálený přístup k souborům přes počítačovou síť.
NIS	Nemocniční informační	

Zkratka	Význam	Ekvivalent / Popis
	systém	
OAIS	Open Archival Information System	Základní norma, z níž vychází většina moderních návrhů na řešení archivů. Archiv zajišťuje dlouhodobé uložení informací od původců a jejich správu, zpřístupňuje obsah uživatelům (badatelům). Systém je „otevřený“ obsahující doporučení, která reagují na neustálý vývoj technologií, formátů, ...
ORP	Obec s rozšířenou působností	ORP jsou mezičlánkem v přenesené působnosti státní správy mezi krajskými a obecními úřady. Obecní úřady obcí s rozšířenou působností mají oproti ostatním obecním úřadům některé oblasti působnosti navíc, a to nejen pro svůj vlastní, základní správní obvod, ale zpravidla i pro další obce v okolí.
PACS	Picture Archiving and Communication System	Síťová a archivační infrastruktura pro ukládání, přenos obrazových informací z medicínských oborů (EKG, X-RAY ....) a jejich zpřístupnění.
PLATTER	PLAnning Tool for Trusted Electronic Repositories	Metodika plánování a údržby důvěryhodných digitálních úložišť
SIP	Submission Information Package	Informační balík obdrženy od původce určený do OAIS. Z něj je vytvořen jeden nebo více archivních informačních balíků (AIP).
SW	Software	
TC	Technologické centrum	Kumulovaná (centralizovaná) hardwarová a softwarová infrastruktura na úrovni kraje nebo ORP, která poskytuje své služby zřizovateli, obcím v jeho správním obvodu a zřízeným nebo založeným organizacím kraje nebo obcí. Viz EgonCentrum
WORM	Write once, read many	Technologie uložení dat, která umožní jednou zapsat informace na ukládací médium a následně zabránit jejich náhodnému nebo záměrnému smazání nebo změně.
ZOOÚ	Zákon o ochraně osobních údajů	

## 7.2.2 Pojmy

Pojem	Význam
-------	--------

Pojem	Význam
Agendový systém	Programový nástroj nebo systém určený pro zpracování údajů odborné agendy.
Archiv	<p>1. Zařízení podle zákona 499/2004 Sb. o archivnictví a spisové službě, které slouží k ukládání archiválií a péči o ně.</p> <p>2. IT systém pro dlouhodobé důvěryhodné ukládání digitálních dokumentů a dat.</p> <p>Především takto užíváno ve spojení "digitální archiv". Použitím termínů "archivace" / "archivní" / "archivační" není na řadě míst v rámci této studie míněna archivace dle zákona 499/2004 o archivnictví a spisové službě, nýbrž čistě technicky pojaté řešení dlouhodobého důvěryhodného ukládání elektronických dokumentů či dat.</p> <p>Hlavním důvodem používání této terminologie je celosvětový "de facto" model dlouhodobého ukládání digitálních dat OAIS (Open Archival Information System), který se do češtiny překládá jako "Otevřený archivní informační systém", z něhož vycházíme při návrhu navrhovaných konkrétních subsystémů KDS, KDR a KDÚ pro Kraj Vysočina.</p>
Archiválie	Takový dokument, který byl vzhledem k době vzniku, obsahu, původu, vnějším znakům a trvalé hodnotě dané politickým, hospodářským, právním, historickým, kulturním, vědeckým nebo informačním významem vybrán ve veřejném zájmu k trvalému uchování a byl vzat do evidence archiválií.
DRAMBORA	Digital Repository Audit Method Based on Risk Assessment. Nástroj pro zpracování vnitřního auditu důvěryhodnosti digitálního úložiště.
eGonCentrum	EGON centrum by mělo být určitým zastřešujícím centrem, které by v sobě mělo slučovat služby poskytované uvnitř úřadu tak také ven směrem o příspěvkovým organizacím a obcím ve správním obvodu obce s rozšířenou působností (kraje). Jedná se o poskytování služeb vzdělávacího centra, administrativního centra a technologického centra.
Elektronická spisová služba	Informační systém, který slouží k zajištění odborné správy dokumentů vzniklých z činnosti původce, popřípadě z činnosti jeho právních předchůdců, zahrnující jejich řádný příjem, evidenci, rozdělování, oběh, vyřizování, vyhotovování, podepisování, odesílání, ukládání a vyřazování ve skartačním řízení, a to včetně kontroly těchto činností.
Langweilův model	Je papírový model historického centra Prahy, který vytvořil v první polovině 19. století Antonín Langweil.
Tier	V tomto významu: vrstvy resp. kategorie datových úložišť (médií), jejichž výběr a použití (vhodnost) pro ukládání různých typů dat se liší na základě úrovně požadované ochrany, výkonnostních požadavků, frekvenci použití dat atd.