

příloha č. 4 – Ideový záměr vybudování Digitálního archivu UK

Ideový záměr vybudování Digitálního archivu UK

1. Současný stav a legislativní předpoklady

1.1. Potřeby UK v oblasti digitální archivace

Univerzita Karlova v Praze (dále UK) je největší českou veřejnou vysokou školou. Je zřizovatelem Ústavu dějin a archivu Univerzity Karlovy (dále ÚDAUK), který je akreditovaným specializovaným archivem podle zákona č. 499/2004 Sb., o archivnictví a spisové službě, v platném znění. UK předává své dokumenty trvalé hodnoty k trvalému uložení do ÚDAUK. Tyto dokumenty se stávají archiváliemi. ÚDAUK má povinnost řádně spravovat a uchovávat svěřené archiválie. V souvislosti s digitalizací vnitřní správy univerzity a elektronizací jejího chodu vznikají dokumenty, které je nutné archiovat stále ve větší míře v digitální podobě. ÚDAUK proto musí zabezpečit řádnou správu a uchování archiválií i v digitální podobě.

1.2. Vznik a správa digitálních dokumentů a digitalizátů

V současné době jsou digitální dokumenty úřední povahy na UK spravovány především v různých elektronických systémech spisové služby (dále ESSS) a v řádově desítkách agendových informačních systémů (dále AIS). Současný stav však není vyhovující, a proto se předpokládá sjednocení správy digitálních dokumentů v nové centralizované spisové službě UK. Dokumenty ze sjednocené spisové služby UK a s ní propojených agendových informačních systémů budou ukládány v centrálním úložišti digitálních dokumentů UK. Z něj budou vyřazovány v rámci skartačního řízení prováděného prostřednictvím rozhraní nové ESSS. Předávané datové balíčky (SIP) budou odpovídat požadavkům Národního standardu pro elektronické spisové služby (dále NSESSS), samotné digitální dokumenty budou ve výstupních formátech uvedených ve vyhlášce č. 259/2012 Sb., o podrobnostech výkonu spisové služby, v platném znění.

Lze předpokládat nárůst objemu dokumentů přebíraných mimo skartační řízení (dokumentů vznikajících mimo úřední činnost UK). Typově půjde o digitální dokumenty předávané v osobních pozůstalostech osob spjatých s UK (elektronická korespondence, osobní digitální archivy a fotoarchivy apod.), dokumenty spolků a vědeckých organizací spjatých s UK a dokumenty shromažďované v rámci vlastní sbírkové a dokumentační činnosti Archivu UK. Tyto dokumenty budou vznikat v řadě různých formátů a při jejich příjmu bude nutná formátová migrace a doplnění nezbytných metadat.

Digitalizáty archiválií, které budou ukládány v Digitálním archivu UK, vznikají v rámci standardizovaného procesu na digitalizační lince Archivu UK. Digitalizovány jsou v zásadě jen archivně zpracované archiválie. Proces digitalizace bude uveden do souladu s doporučeními uvedenými v Metodice pro vytváření bezpečnostních kopií archiválií v digitální podobě (Národní archiv, Státní oblastní archiv v Praze, Praha 2015, verze 1.0). V minulosti vznikaly některé digitalizáty v jiných

příloha č. 4 – Ideový záměr vybudování Digitálního archivu UK

standardech, metadata těchto digitalizátů bude nutné při příjmu do Digitálního archivu UK konvertovat.

1.3. Situace Archivu UK a legislativní možnosti ukládání digitálních archiválií

Aktuální stav uložení a správy archiválií v digitální podobě a digitalizátů neodpovídá požadavkům zákona č. 499/2004 Sb. Archiválie jsou ukládány a spravovány v diskovém úložišti, v úložišti CESNET a na fyzických nosičích. Archiválie v digitální podobě nejsou předávány do Národního archivu ani do jiného akreditovaného archivu, kterému by bylo uděleno oprávnění k ukládání archiválií v digitální podobě.

§ 15 odst. 3 zákona č. 499/2004 Sb. umožňuje tři varianty řešení problematiky ukládání digitálních archiválií:

1) Archiválie v digitální podobě náležející do péče Archivu UK se uloží v Národním archivu (Národním digitálním archivu, který je jeho součástí). UK musí zajistit předání archiválií do Národního archivu ve struktuře odpovídající SIP balíčku požadovanému Národním digitálním archivem. Realizace projektu Národního digitálního archivu má několikaleté zpoždění, stále funguje jen v poloprovozním režimu a rozsah jeho funkcionalit bude oproti původním deklaracím patrně omezen. Předáním digitálních archiválií a digitalizátů se Archiv UK nezbaví povinnosti správy těchto archiválií.

2) Archiválie v digitální podobě náležející do péče Archivu UK se na základě písemné dohody zřizovatelů archivů uloží v archivu, jehož zřizovateli je uděleno oprávnění k ukládání archiválií v digitální podobě. UK musí zajistit předání archiválií ve struktuře odpovídající SIP balíčku požadovanému Národním digitálním archivem. V ČR zatím nebylo žádnému archivu uděleno povolení ukládat archiválie v digitální podobě podle § 60a zákona č. 499/2004 Sb.

3) ÚDAUK získá oprávnění k ukládání archiválií v digitální podobě ve smyslu § 60a zákona č. 499/2004 Sb. Navržený systém musí splnit požadavky zákona č. 499/2004 Sb., v jeho platném znění včetně prováděcích norem a metodik vydávaných Národním archivem a Ministerstvem vnitra.

2. Potřeba a možnosti vybudování Digitálního archivu UK

2.1. Zdůvodnění vhodnosti vybudovat vlastní Digitální archiv UK

Správou digitálních archiválií UK musí být dle zákona č. 499/2004 Sb. pověřen její specializovaný archiv – ÚDAUK. Proto bude Digitální archiv UK začleněn do ÚDAUK. Vybudování Digitálního archivu UK umožní ÚDAUK získat oprávnění k ukládání archiválií v digitální podobě dle § 60a zákona č. 499/2004 Sb. Vybudování vlastního digitálního archivu umožní UK zabezpečit prostřednictvím svého specializovaného archivu maximální možnou úroveň a nezávislost v péči o své digitální archiválie a digitalizáty analogových archiválií. Vybudování vlastního digitálního archivu také umožní řádně využívat uložené archiválie pro správní účely a pro vědecký výzkum realizovaný na UK. Vlastní řešení umožní maximální účelnou vazbu na specifické informační systémy UK a správu specifických univerzitních digitálních dokumentů a metadat. Vlastním digitálním archivem UK navíc

příloha č. 4 – Ideový záměr vybudování Digitálního archivu UK

získá nezávislost na problematickém a stále odkládaném projektu státního Národního digitálního archivu. UK při vybudování digitálního archivu zúročí odborné kapacity ÚDAUK a Ústavu výpočetní techniky UK (dále ÚVT UK), který bude poskytovat technickou podporu při jeho realizaci a provozu.

2.2. Dostupné technické zázemí (univerzitní hardware a využitelný software)

Pro hlavní a pracovní úložiště může být využito zázemí dvou datových center v Praze (v lokalitě Praha 1 /Rektorát UK/ a Praha – Jinonice). Obě lokality jsou připojené do optické sítě Pasnetu, disponují dostatečným chladicím výkonem a příkonem elektrické energie, který je zálohován dieselagregáty. Jako úložiště slouží dvě vzájemně zálohovaná disková pole (každé v jednom datovém centru), která je možné rozšířit dokoupením disků. Dále jsou k dispozici páskové zálohovací jednotky. Je implementována virtualizační infrastruktura postavená z části na VMware a z části na Oracle VM server, v plánu je instalace cloudového rozhraní (pravděpodobně OpenStack).

Pro záložní pracoviště je možné uvažovat o využití datových center při mimopražských fakultách UK v Plzni a Hradci Králové.

2.3. Dostupné „lidské zdroje“

Archiv UK v současné době disponuje odborníky s kvalifikací pro péči o uložené archiválie. Nedisponuje odborníky pro správu aplikací digitálního archivu a dalšími potřebnými informačními specialisty.

ÚVT UK je schopen zajistit administraci serverů včetně operačních systémů a zálohování aplikací a souborů archivu. Pro správu případných aplikací nemá v současnosti ÚVT UK kapacity.

Specifikace potřeby nových odborníků pro provoz Digitálního archivu UK bude upřesněna ve studii proveditelnosti Digitálního archivu UK.

2.4. Umístění úložišť

Digitální archiv musí dle platné legislativy disponovat nejméně hlavním a záložním úložištěm pro uložení archiválií. Vedle toho by měl disponovat i jedním úložištěm pracovním.

Umístění a zabezpečení úložišť musí respektovat a splňovat podmínky uvedené v aktuálním znění zákona č. 499/2004 Sb., o archivnictví a spisové službě, jeho prováděcích vyhláškách a v příslušných metodikách Odboru archivní správy a spisové služby Ministerstva vnitra ČR a Národního archivu ČR. Umístění úložišť musí splňovat podmínky dané pro akreditaci ÚDAUK jako specializovaného archivu (zejména § 58 a § 61 zákona č. 499/2004 Sb.) a podmínky nezbytné pro získání oprávnění pro ukládání archiválií v digitální podobě (zejména § 60a a § 61, odst. 2 a 4 zákona č. 499/2004 Sb.). Hlavní a záložní úložiště musí být umístěno v lokalitách vzdálených minimálně 50 km vzdušnou čarou.

Prostory úložišť a úložiště samotná musí být ve správě UK a navazovat na informační infrastrukturu UK. UK proto preferuje, aby úložiště byla umístěna v objektech, které UK již nyní vlastní nebo které se stávajícími objekty UK funkčně souvisí. Lokality pro umístění úložišť musí umožňovat

příloha č. 4 – Ideový záměr vybudování Digitálního archivu UK

vybavení nezbytnými energetickými sítěmi (včetně jejich zálohování pro případ výpadku), kapacitním datovým připojením a dalšími technologiemi nezbytnými pro provoz datového úložiště (klimatizace apod.). Lokality pro umístění úložišť musí disponovat dostatečnými prostorovými a technickými rezervami pro umístění dodatečných technologií pro rozšíření datové kapacity úložišť (nejméně v rozsahu 100 % oproti původně pořizované kapacitě).

3. Očekávané řešení

3.1. Legislativa, normy a závazné metodiky

Oblast archivnictví, výběr archiválií (v analogové i digitální podobě), jejich ukládání, péči o ně a zpřístupnění komplexně upravuje **zákon č. 499/2004 Sb., o archivnictví a spisové službě a o změně některých zákonů**, v platném znění. Možnost vybudování digitálního archivu (získání oprávnění pro ukládání archiválií v digitální podobě) upravuje § 60a tohoto zákona. Při vybudování digitálního archivu je však nutné respektovat zákon jako celek. Zákon č. 499/2004 Sb. doplňují jeho prováděcí **vyhláška č. 645/2004 Sb., kterou se provádějí některá ustanovení zákona o archivnictví a spisové službě a o změně některých zákonů**, v platném znění, a **vyhláška č. 259/2012 Sb., o podrobnostech výkonu spisové služby**, v platném znění.

Řešení Digitálního archivu UK musí splňovat nároky uvedené ve **Vzorovém provozním řádu archivu oprávněného k ukládání archiválií v digitální podobě**, který byl zveřejněn ve Věstníku ministra vnitra, částka 65/2012 (část II). Vzhledem k tomu, že Digitální archiv UK bude součástí Archivu UK, který je akreditovaným archivem podle zákona č. 499/2004 Sb., bude nutné zohlednit i příslušné pasáže **Pokynu ředitele odboru archivní správy MV, kterým se specifikují podklady pro rozhodnutí o akreditaci nebo pro prokázání splnění podmínek stanovených v § 61 zákona č. 499/2004 Sb. (čj. AS - 113/Ř-2005)**. Funkcionalita Digitálního archivu pro zpracování archiválií musí respektovat platná **Základní pravidla pro zpracování archiválií** (aktuální II. opravené vydání z roku 2015). Provoz Digitálního archivu UK může ovlivňovat i zákon č. 181/2014 Sb., o kybernetické bezpečnosti a změně některých zákonů, a jeho prováděcí vyhlášky (č. 316/2014 Sb. a č. 317/2014 Sb.). Dopad tohoto zákona na oblast archivnictví však není podle dosavadních výkladů zcela jasný a musí být určen dalším právním výkladem.

Řešení Digitálního archivu UK musí odpovídat principům uvedeným v normě **ČSN ISO 14721 (319620) - Systémy pro přenos dat a informací z kosmického prostoru - Otevřený archivační informační systém - Referenční model** a zohlednit možný budoucí audit podle normy **ČSN ISO 16363 (319621) - Systémy pro přenos dat a informací z kosmického prostoru - Audit a certifikace důvěryhodných digitálních úložišť**.

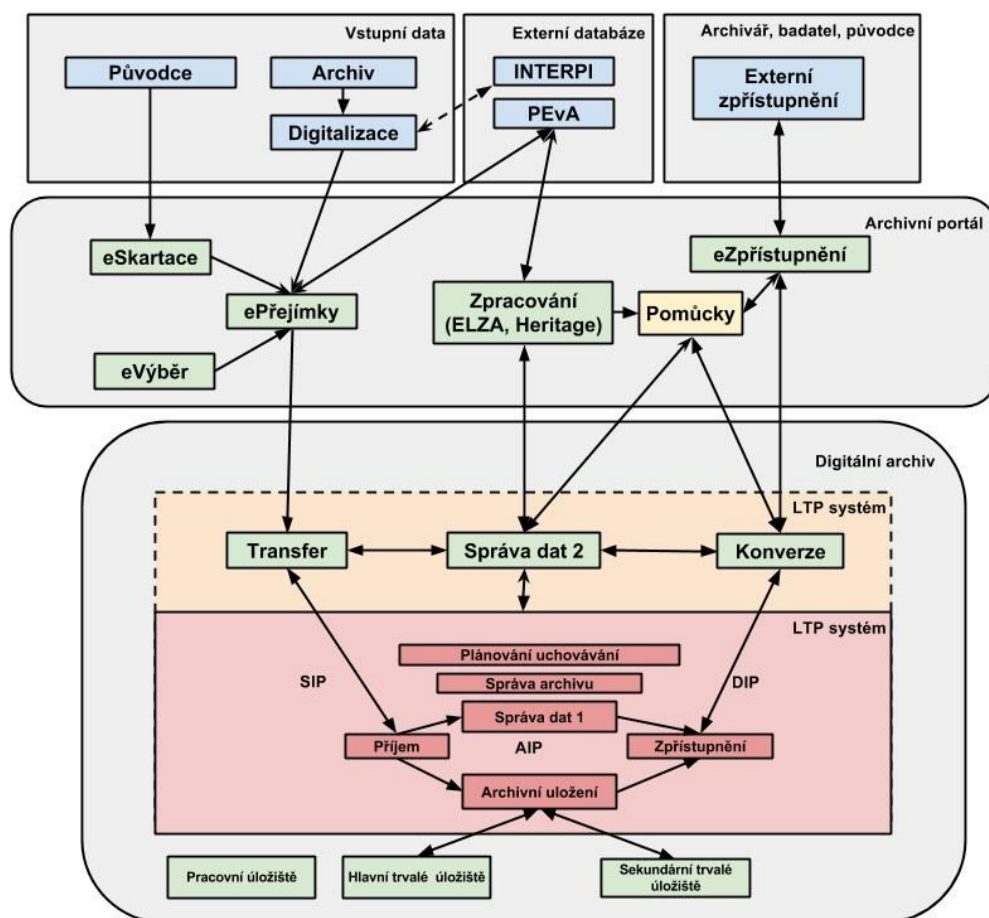
Legislativa a metodika týkající se archivnictví prochází velmi častými novelizacemi, je proto nutné sledovat aktuální vývoj a zohlednit novelizace a nově vznikající předpisy, standardy a metodiky.

příloha č. 4 – Ideový záměr vybudování Digitálního archivu UK

3.2. Očekávané funkcionality, očekávaná kapacita úložišť

Digitální archiv UK umožní příjem, správu, uložení, dlouhodobou ochranu a zpřístupnění digitálních archiválií. Součástí digitálního archivu bude archivní portál, systém pro dlouhodobou ochranu archiválií a samotná datová úložiště.

Navrhované funkční vztahy jednotlivých funkcionalit digitálního archivu jsou zobrazeny v připojeném schématu.



Obrázek 1. Schéma funkčních celků digitálního archivu UK

3.2.1. Slovníček ke schématu

AIP je archivní informační balíček obsahující veškerá ochraňovaná data a metadata podle požadavků informačního modelu OAIS (Open Archival Information System dle ČSN ISO 14721).

Archiv je odborné pracoviště sloužící pro ukládání archiválií a k všestranné péči o ně.

Archivní pomůcka je informační systém, který se vytváří při archivním zpracování a slouží pro evidenci a orientaci v obsahu a časovém rozsahu archivního fondu.

příloha č. 4 – Ideový záměr vybudování Digitálního archivu UK

Archivní uložení je funkční celek archivního uložení dle OAIS (archival storage functional entity), který zahrnuje služby a funkce využívané k ukládání a získávání archivních informačních balíčků.

Digitalizace je převod analogového dokumentu do digitální (numerické) formy.

DIP je balíček ve formátu pro Zpřístupnění, vytváří se konverzí balíčku AIP.

ELZA je aplikace pro zpracování archiválií a tvorbu archivních pomůcek vyvíjená Odborem archivní správy a spisové služby MV ČR.

ePřejímky je modul, který umožňuje příjem digitálních dokumentů k trvalému uložení v digitálním archivu (vybraných prostřednictvím nástrojů eSkartace, eVýběr a z digitalizačních linek).

eSkartace je nástroj pro provádění skartačních řízení z elektronických systémů spisových služeb.

eVýběr je nástroj pro výběr digitálních dokumentů mimo skartační řízení (dokumentů neúřední povahy).

Externí zpřístupnění jsou webové portály mimo samotný digitální archiv sloužící pro zpřístupnění archiválií v digitální podobě a digitalizátů (např. <http://www.apex-project.eu/index.php/en/>).

eZpřístupnění je modul, který umožňuje zpřístupnění archiválií v digitální podobě a digitalizátů.

Heritage je aplikace pro zpracování archiválií a tvorbu archivních pomůcek vyvíjená firmou LemonBone s. r. o.

Hlavní trvalé úložiště je prostor pro ukládání AIP balíčků včetně jejich metadat.

INTERPI (Interoperabilita v paměťových institucích) je znalostní databáze paměťových institucí fungující na bázi národních autorit. <http://www.interpi.cz/>

Konverze je komponenta pro přípravu dat do formátu, ve kterém je očekávají koncoví uživatelé nebo další systémy. V některých řešeních může být integrována s komponentou Zpřístupnění.

LTP systém je subsystém odpovídající požadavkům referenčního modelu OAIS. Ve schématu jsou znázorněny dvě varianty rozsahu tohoto subsystému. V případě použití standardního open-source nebo komerčního řešení předpokládáme, že bude třeba implementovat komponenty Transfer, Konverze a Správa dat 2 s funkcemi nad rámec funkcí popsaných v OAIS a reálně dostupných ve standardním systému. V případě vlastního vývoje LTP lze předpokládat, že Transfer bude spojen s komponentou Příjem, Konverze s komponentou Zpřístupnění a Správa dat bude jen jedna nad všemi typy dat a balíčků (tedy nebude rozdělena na části Správa dat 1 a Správa dat 2).

PEvA (Program pro evidenci archiválií) je celostátně užívaný software, který slouží k evidenci tzv. Národního archivního dědictví.

Plánování uchovávání je funkční celek plánování uchovávání dle OAIS (preservation planning functional entity), který poskytuje služby a funkce pro sledování okolí archivu OAIS a který stanovuje doporučení a plány uchovávání k zajištění dlouhodobé přístupnosti a srozumitelnosti a dostatečné využitelnosti informací uložených v archivu OAIS pro určenou skupinu, a to i v případě zastarání původního počítačového prostředí.

Pracovní úložiště je diskový prostor, kde LTP systém ukládá a zpracovává balíčky SIP a DIP, případně kde jsou umístěna data během transferu, nebo aktualizace v průběhu zpracování komponentou Správa dat.

příloha č. 4 – Ideový záměr vybudování Digitálního archivu UK

Příjem je funkční celek příjmu dle OAIS (ingest functional entity), který zahrnuje služby a funkce, jež od tvůrců přijímají vstupní informační balíčky, připravují archivní informační balíčky určené k uložení a zajišťují, aby archivní informační balíčky a k nim náležející podpůrné popisné informace byly zařazeny do archivu OAIS.

Původce je každý, z jehož činnosti dokument vznikl. Za dokument vzniklý z činnosti původce se považuje rovněž dokument, který byl původci doručen nebo jinak předán.

Sekundární trvalé úložiště je prostor pro ukládání balíčků AIP ve druhé lokalitě splňující požadavky zákona (50km od Hlavního trvalého úložiště s jinou technologií ukládání dat atd.).

SIP je vstupní informační balíček ve formátu, ze kterého je Příjem schopen vytvořit balíček AIP.

Správa archivu je funkční celek správy dle OAIS (administration functional entity), který zahrnuje služby a funkce potřebné pro řízení běžného provozu ostatních funkčních celků archivu OAIS.

Správa dat 1 je funkční celek správy dat dle OAIS (data management functional entity), který zahrnuje služby a funkce pro ukládání, údržbu a zpřístupnění různorodých informací. Příklady těchto informací jsou katalogy a seznamy položek, které lze získat z funkčního celku archivního uložení, algoritmy, pomocí kterých mohou být zpracována získaná data, statistické údaje o přístupech koncových uživatelů, faktury pro koncové uživatele, rámcové objednávky, bezpečnostní opatření a harmonogramy, pravidla a postupy archivu OAIS.

Správa dat 2 je komponenta na stejné úrovni, jako jsou komponenty Transfer a Konverze. Některá standardní řešení pro LTP systémy (např. open-source systém Archivematica) disponují v jádru LTP systému komponentou Správa dat s omezenou funkcionalitou. Architektury využívající tyto systémy pak musí pro praktické použití implementovat další komponentu Správa dat ještě vně LTP systému, pomocí níž jsou indexována veškerá metadata včetně technických metadat a kde jsou implementovány funkce, které v jádru LTP systému chybí.

Transfer je komponenta pro přípravu dat pro uložení do LTP systému. Umožňuje konverzi z dodavatelského balíčku do formátu SIP, se kterým se pracuje v Příjmu.

Zpřístupnění je funkční celek zpřístupnění dle OAIS (access functional entity), který zahrnuje služby a funkce, jež činí archivované informační jednotky a příbuzné služby dostupnými pro koncové uživatele.

3.2.2. Popis funkcionalit jednotlivých funkčních celků

3.2.2.1. Archivní portál

Archivní portál umožní příjem, zpracování a zpřístupnění digitálních archiválií, replik digitálních dokumentů a digitalizátů.

V oblasti příjmu umožní příjem dokumentů ve skartačním řízení. Bude obsahovat funkční rozhraní pro příjem SIP balíčků s úředními dokumenty produkovány systémy spisových služeb a dalšími informačními systémy, které odpovídají Národnímu standardu pro systémy elektronických spisových služeb. Archivní portál bude obsahovat funkční rozhraní pro provádění výběru dokumentů ve skartačním řízení (schéma **eSkartace** → **ePřejímky**). Zároveň bude schopen komunikovat s aplikacemi používaných informačních systémů a Národního archivu pro provádění skartačního řízení. Archivní portál umožní také příjem digitálních dokumentů a replik digitálních archiválií mimo skartační řízení, přičemž umožní tvorbu příslušných SIP balíčků odpovídajících české legislativě a mezinárodním ISO normám (schéma **eVýběr** → **ePřejímky**). Archivní portál dále umožní příjem

příloha č. 4 – Ideový záměr vybudování Digitálního archivu UK

digitalizátů archiválií a příslušných metadat (v podobě popsané v bodě 1.2.) produkovaných digitalizačními linkami a umožní tvorbu příslušných SIP balíčků (schéma **Digitalizace → ePřejímky**).

Archivní portál umožní příjem replik digitálních archiválií, celých archivních fondů nebo jejich částí včetně jejich metadat a standardizovaných archivních pomůcek z jiných digitálních archivů (archivů oprávněných podle zákona č. 499/2004 Sb. ukládat archiválie v digitální podobě) v rámci delimitací mezi jednotlivými archivy. Ve všech případech bude možný příjem prostřednictvím webového rozhraní, FTP, NFS, prostřednictvím fyzických nosičů digitálních dat, ze kterých budou data nahrávána prostřednictvím pracovní stanice archiváře, a v případě výstupů z digitalizačních linek i prostřednictvím standardizovaného automatizovaného rozhraní. Při příjmu dokumentů archivní portál umožní aktivní výměnu metadat s databázemi pro správu informací o archivních fondech a datech používaných v paměťových institucích (zejména s Programem pro evidenci archiválií a databází pro Interoperabilitu v paměťových institucích, schéma **Externí databáze – INTERPI a PEvA**).

V oblasti správy archiválií umožní archivní portál archivní zpracování archiválií a tvorbu archivních pomůcek v souladu s platnými pravidly pro zpracování archiválií vydávanými Odborem archivní správy a spisové služby MV ČR. ÚDAUK jako budoucí provozovatel digitálního archivu preferuje využít obecných systémů používaných v českých archivech (schéma **Zpracování**) (např. systém Heritage, systém ELZA vyvíjený Odborem archivní správy a spisové služby MV ČR) a jejich integraci do budoucího řešení digitálního archivu UK. Použitý systém musí umožnit zpracování kompletních archivních fondů tvořených analogovými i digitálními dokumenty. Metadata vzniklá archivním zpracováním bude systém předávat rozhraní pro správu dat v archivním úložišti (budou zapisovány do AIP balíčků digitálních dokumentů) a do databáze spravované v Programu pro evidenci archiválií (PEvA). Vzniklá metadata budou formou archivních pomůcek (schéma **Pomůcky**) a rozhraní pro zpřístupnění archiválií přístupná uživatelům archivu. Archivní pomůcky budou ukládány jako samostatné AIP balíčky, systém umožní jejich předávání jiným digitálním archivům a jejich zpřístupnění prostřednictvím jiných archivních portálů (zejména Národního archivního portálu). Archivní portál také umožní využití archivních pomůcek v rámci digitalizace.

V oblasti zpřístupnění archiválií archivní portál umožní zpřístupnění uložených archiválií ve formě DIP balíčků (shodných s uloženými digitálními archiváliemi a digitalizáty, či pro potřeby zpřístupnění upravených). Systém umožní různý uživatelský přístup (schéma **eZpřístupnění**) daný rolí přístupovatele (administrátor systému, archivář – správce digitálního obsahu, zástupce původce, badatel). Systém umožní přístup ke zpracovaným archiváliím prostřednictvím archivních pomůcek a případných specializovaných archivních databází. Zároveň umožní zpřístupnění nezpracovaných archiválií vybraným rolím přístupovatelů na základě dostupných metadat, případně na základě výběru určeného uživatelem s vyšší přístupovou rolí. Systém umožní zpřístupnění prostřednictvím webového rozhraní integrovaného do archivního portálu, místního intranetového rozhraní digitálního archivu a ve vybraných případech i prostřednictvím externích nástrojů pro dálkové zpřístupnění archiválií (systém musí být schopen vydávat standardizované DIP balíčky externím systémům, schéma **Externí zpřístupnění**). Systém umožní výdej replik digitálních archiválií, celých archivních

příloha č. 4 – Ideový záměr vybudování Digitálního archivu UK

fondů, archivních pomůcek nebo jejich částí jiným digitálním archivům. Tento výdej bude možný prostřednictvím webového rozhraní nebo prostřednictvím fyzických datových nosičů.

3.2.2.2. Systém pro dlouhodobé uložení a ochranu digitálních archiválií

Součástí řešení digitálního archivu bude systém pro dlouhodobé uchovávání digitálních dat (schéma **LTP systém**). Jeho architektura, informační model a funkční model budou vycházet z referenčního modelu OAIS (ČSN ISO 14721) a metadata informačních **balíčků AIP** budou v kontejnerovém XML formátu METS (Metadata Encoding & Transmission Standard, viz <http://www.loc.gov/standards/mets/>), s vnořenými technickými, administrativními, ochrannými a popisnými metadaty. Systém by měl respektovat standard PREMIS (viz <http://www.loc.gov/standards/premis/>) pro archivační/ochranná metadata dle OAIS. Specifickým typem AIP balíčku budou informační balíčky AIP obsahující archivní pomůcky (schéma **Pomůcky**), které budou spravovány externím systémem a v LTP systému budou také předmětem ochrany.

Systém by měl mít otevřená rozhraní a měl by být kastomizovatelný tak, aby byl využitelný i pro jiná data než pro data z digitálního archivu (data knihoven apod.). Systém by měl mít otevřená rozhraní umožňující rozšíření o další nástroje (např. pro validaci, charakterizaci a konverze formátů), jejichž doplnění mohou být požadována jak ve výchozí instalaci, tak i později v průběhu provozu systému a to nejen ve své počáteční instalaci, ale i průběžně v závislosti na vývoji daných nástrojů.

Jádrem LTP systému bude systém pro fyzickou ochranu bitů tvořících AIP balíčky, tedy systém pro archivní uložení (schéma **Archivní uložení**). Zde budou uložena data určená pro trvalé uložení v podobě AIP balíčků. Systém bude dále disponovat databází a indexem metadat AIP balíčků (schéma **Správa dat 1** v LTP systému). Součástí systémů digitálního archivu může být i samostatný modul (index a databáze) aktuálních metadat archivních balíčků AIP a rozpracovaných balíčků SIP (schéma **Správa dat 2**). Entity **Transfer** a **Konverze** zajistí přípravu dat pro LTP systém a jejich konverzi podle potřeb systémů využívajících balíček DIP i nad rámec popisu referenčního modelu OAIS a budou využívat také vlastní pracovní úložiště. Zpracování musí probíhat jednorázově, manuálně nebo automaticky nad dávkami balíčků. *(V závislosti na použitém LTP systému mohou být komponenty **Transfer**, **Konverze** a **Správa dat 2** více či méně integrovány s entitami **Příjem**, **Zpřístupnění** a **Správa dat 1**).*

Funkční entita **Příjem** v souladu s požadavky vyplývajícími z referenčního modelu OAIS provede kontroly kvality dodaných balíčků a jejich konverzi do formátu AIP balíčku LTP systému. Zpracování vstupu dat v Příjmu by mělo být řízeno konfigurovatelným workflow systémem (např. Business proces management system), který umožní administrátorovi nastavit jaké kontroly, s jakým nastavením a jakými nástroji bude systém pro danou skupinu dat provádět. Zde by měla být validována metadata a doplněna administrativní, technická a ochranná metadata LTP systému (identifikátory, záznam informací souvisejících s příjmem dokumentů, identifikace a validace formátů, extrakce technických metadat z digitálních objektů). Komponenta **Příjem LTP systému** bude integrovat systémy třetích stran pro identifikaci formátů, validaci formátů a extrakci technických metadat. Zároveň se předpokládá, že se tyto nástroje budou časem měnit a operace validace formátů bude třeba v budoucnosti opakovat za použití jiných nástrojů. Validní balíčky AIP budou uloženy do

příloha č. 4 – Ideový záměr vybudování Digitálního archivu UK

Archivního úložiště (kde budou dále replikovány do více lokalit a pravidelně kontrolovány v souladu s nastavitelnou politikou) a zároveň bude zaznamenána informace o uložení a metadata do funkční komponenty **Správa dat 1**. Ta bude indexovat všechna metadata balíčků AIP a poskytne pokročilé vyhledávací rozhraní, které umožní také uložení vyhledávacích dotazů a jejich využití v dalších částech LTP systému (schéma **Plánování uchovávání, Zpřístupnění**).

Entita **Zpřístupnění** LTP systému v souladu s OAIS na základě dotazu ze Správy dat 1 provede dodání balíčku DIP, který může být následně v entitě **Konverze** upraven do podoby požadované uživatelem nebo dále zpřístupněn uživateli s využitím externích prohlížečů (konverze metadat a formátové konverze, případně pokročilé funkce pro AV dokumenty, streamování apod.).

Implementace funkcí entity **Plánování uchovávání** přímo v LTP systému by měla minimálně umožnit vyhodnocení rizik spojených s jednotlivými souborovými formáty, disponovat aktualizovatelnou databází formátů, rizik a konverzních nástrojů a umožnit nastavení a realizaci akcí vyplývajících z plánu dlouhodobého uchovávání, tj. především provedení transformace a doplnění metadat a formátové migrace. Všechny změny balíčků AIP uskutečněné v Plánování uchovávání budou zaznamenány do metadat AIP balíčků a budou vytvořeny nové verze AIP balíčků. Při zpracování balíčků používá entita Plánování uchovávání vždy kopie balíčků AIP exportované z pracovního úložiště. V žádném případě nepřistupuje přímo k AIP balíčkům v archivním uložení, ale v souladu s OAIS využívá mechanismy entity Zpřístupnění k jejich získání a mechanismy Příjmu k jejich opětovnému vložení do Archivního úložiště (jako nové verze). Část funkcí popsaných v OAIS entitě **Plánování uchovávání** (jako sledování vývoje technologií, sledování potřeb určené skupiny, definice standardů a politik) jsou spíše procesní funkce, kde nepředpokládáme plnou implementaci a podporu těchto funkcí budovaným informačním systémem digitálního archivu.

Funkce entity **Správa archivu** a další základní služby poskytují prostředí pro provoz, konfiguraci a řízení LTP systému. Předpokládáme, že LTP systém bude muset disponovat především mechanismy pro správu uživatelů a jejich oprávnění, jejich autentizaci, umožní konfiguraci systému, řízení a logování procesů a jeho workflow, a zapojení externích nástrojů. Systém by měl být dostatečně rozšiřitelný tak, aby umožnil zapojení nových nástrojů a zásuvných modulů pro rozšiřování funkcionality. Workflow systém (BPM) by měl vedle konfigurace kroků workflow umožnit také definování sady balíčků SIP nebo AIP, nad kterými by měly akce probíhat, a také časové nastavení okamžiku spuštění nějaké operace (*scheduling*).

Všechny komponenty LTP systému budou disponovat v souladu s OAIS nástroji pro reportování, které usnadní administraci LTP systému. Systém by měl minimálně obsahovat reporty o výkonu zpracování dat, reporty o obsahu trvalého úložiště, objemech, formátech dat a rizicích s nimi spojených, reporty o zpřístupnění, audit reporty.

3.2.2.3. Archivní úložiště

Systém pro dlouhodobé uložení a ochranu archiválií bude pro zajištění bitové ochrany využívat technické prostředky archivního úložiště. Musí být nezávislý na technologii uložení dat a vzhledem k dalším systémům archivu musí řešení pro archivní uložení poskytovat standardní POSIX II rozhraní. Technologie použité v archivním úložišti by tak měly být pro další systémy transparentní

příloha č. 4 – Ideový záměr vybudování Digitálního archivu UK

a mělo by být možné použít v budoucnu jakékoli technologie ukládání dat. Vlastní fyzické uložení dat by mělo být dostatečně dokumentováno a provozovatel digitálního archivu by měl mít možnost ovlivnit způsob uložení dat (uspořádání dat do složek; způsob zabalení dat do kontejneru zip/tar, bagit; fyzický vs logický AIP balíček).

AIP balíčky budou obsahovat všechna data a metadata tak, aby bylo možné získat kompletní balíček AIP i bez dalších systémů archivu přímo z úložiště.

Úložiště by mělo v souladu s požadavky norem a procesů archivu umožňovat aktualizaci metadata a verzování balíčků a transformaci jejich obsahu při realizaci plánů dlouhodobé ochrany s tím, že události spojené s AIP balíčky budou logovány a zaznamenány v jejich metadatach.

Vedle úložiště pro trvalé uložení AIP balíčků bude systém pro dlouhodobé uložení potřebovat pracovní úložiště, kde budou data dočasně ukládána při zpracování vstupu dat, zpřístupňována a konvertována při výstupu z digitálního archivu, a ukládána při transformacích a kontrolách prováděných administrátorem LTP systému pro účely plánování uchovávání.

Základním požadavkem na budování digitálního archivu vyplývajícím z legislativy je použití více technologií ukládání dat a uložení AIP balíčků ve více kopiích na více místech. Systémy archivního úložiště by měly disponovat nástroji pro zajištění replikace nebo zrcadlení AIP balíčků, nastavením frekvence kontrol integrity, mechanismy pro porovnávání instancí jednoho AIP ve více lokacích. Také by měly umožnit kontroly integrity (*fixity checks*) s možností nastavení frekvence kontrol pro definovanou skupinu balíčků.

Kapacita hlavního a záložního úložiště je plánována na uložení dokumentů přebíraných v období 15 let od zprovoznění digitálního archivu.

	Současný stav	Roční přírůstek	Za 15 let
Dokumenty z ESSS a AIS	-	90 – 120 GB	2 TB
Kvalifikační práce	0,8 TB	80 GB	2 TB
Digitalizáty	8 TB	4 TB	66 TB
Mimoskartační přírůstky (pozůstalosti)	-	2 TB	30 TB
Celkem			100 TB

Při zprovoznění Digitálního archivu UK je očekávaná kapacita **50 TB**, zbylých 50 TB z plánované kapacity může být doplněno podle aktuální potřeby. Použitá technologie musí umožnit rozšíření kapacity do 6 měsíců od vznesení požadavku.

Pracovní úložiště by mělo disponovat cca poloviční kapacitou, tedy 50 TB.

3.2.2.4. Integrace a rozhraní s externími systémy

- Národní archiv – možnost exportu balíčků ve struktuře vyžadované Národním digitálním archivem.
- Výměna balíčků s jinými archivy – ve struktuře vyžadované Národním digitálním archivem.

příloha č. 4 – Ideový záměr vybudování Digitálního archivu UK

- Možnost přenosu informací ze systémů PEvA a INTERPI – v rozsahu umožněném rozhraními těchto systémů.
- Automatizované dávkové zpřístupnění metadat pro externí systémy pro zpřístupnění archiválií (APEX, Europeana apod.), a to na základě standardního komunikačního protokolu (např. OAI-PMH).
- Příjem SIP balíčků ze systémů spisových služeb dle NSESSS.
- Výdej DIP balíčků ve standardu uvedeném v NSESSS.
- Integrace nástrojů pro validaci a charakterizaci formátů.
- Integrace externích formátových registrů.
- Integrace konverzních nástrojů.

3.3. Realizovatelnost, očekávaný harmonogram

Digitální archiv UK musí být realizovatelný a provozuschopný z hlediska investiční, finanční a personální náročnosti budoucího provozu. Samotnou investici bude nutné vzhledem k odhadované výši a rozpočtovým možnostem UK financovat z mimorozpočtových zdrojů. V rámci přípravy proto bude nutné prověřit dostupnost a využitelnost těchto zdrojů. Realizace digitálního archivu bude muset respektovat reálnou situaci v oblasti dostupných vlastních zdrojů. ÚVT UK nebude moci poskytnout rozsáhlejší kapacity pro vlastní vývoj, omezené budou i dostupné kapacity pro samotnou implementaci.

Realizovatelnost Digitálního archivu UK ovlivní neustálá legislativa a standardy v oblasti archivnictví a spisové služby. Přípravu Digitálního archivu UK bude nutné průběžně konzultovat s příslušnými státními orgány (zejména s Odborem archivní správy a spisové služby MV ČR a Národním archivem ČR), bude nutné se v předstihu připravit i na dosud legislativně nevyžadované požadavky.

S riziky bude spojena i samotná realizace Digitálního archivu UK, protože dodávky technologií a služeb musí UK realizovat formou výběrových řízení v režimu zákona č. 137/2006 Sb., o veřejných zakázkách, což přinese zvýšené nároky na právní služby při přípravě a samotné realizaci a může ovlivnit i harmonogram realizace.

Z časového hlediska musí realizace respektovat postupně vzrůstající potřebu UK v oblasti digitálního archivace. Jako aktuální se nyní jeví potřeba archivovat digitalizáty analogových archiválií uložených v Archivu UK, tento požadavek je možné po omezenou dobu realizovat dočasnými náhradními řešeními. Podobně je možné postupovat i u dokumentů vybíraných v mimoskartačním řízení. Do spisových služeb na UK se v současné době dostávají první digitální dokumenty, které bude nutné po uplynutí skartační lhůty archivovat. S ohledem na nastavení skartačních lhůt a životnost používaných dočasných řešení je proto nutné vybudovat digitální archiv **nejpozději do roku 2021**.

3.4. Provozní požadavky

Základním personálním provozním požadavkem pro vybudování Digitálního archivu UK je omezení požadavků na nové odborné zaměstnance, tedy maximální využití stávajících personálních

příloha č. 4 – Ideový záměr vybudování Digitálního archivu UK

zdrojů. Je potřebné definovat i případné skryté nedostatky v procedurách archivu, jejichž náprava bude vyžadovat další podstatné a nákladné organizační změny.

Při volbě použitých technologií a systémů je třeba respektovat požadavek na ekonomičnost provozu Digitálního archivu UK, který bude (na rozdíl od samotné investice) financován z běžného rozpočtu UK. Výběr technologií a systémů musí také respektovat omezené možnosti na investiční obnovu (zvolené technologie a systémy musí tedy mít dostatečnou dobu životnosti bez nutnosti dalších investic).

3.5. Požadavky rozvoje a udržitelnosti (možnost rozšíření, životnost pořízeného řešení)

Prostory navržené pro úložiště musí disponovat dostatečnou kapacitou, anebo být snadno rozšiřitelné tak, aby umožňovaly, s nutnými úpravami a rozšířeními technologií, provoz Digitálního archivu UK po dobu nejméně 30 let od jeho zprovoznění.

Technologie použité pro úložiště digitálního archivu by měly navazovat na technologie již použité na UK ve zvolené lokalitě, a to jak z hlediska jejich technického charakteru, tak z hlediska odbornosti a zkušenosti pracovníků univerzity.

U použitých hardwarových technologií musí být zaručena jejich životnost a zachování jejich dostupnosti, a to minimálně po dobu 10 let od jejich nasazení do provozu. Stejně tak musí být po tuto dobu zaručena i rozšiřitelnost řešení s ohledem na plánovaný nárůst objemu dat s dostatečnou rezervou pro případ růstu objemu dat nad očekávaný objem. Toto se týká jak kapacit úložiště, tak přenosových a výpočetních kapacit. Dodavatel musí minimálně na tuto dobu zaručit i podporu použitých technologií.

Lokality zvolené pro úložiště musí umožňovat rozšíření objemu dodávek elektrické energie, včetně nouzového zabezpečení pro případ výpadku dodávky z veřejné sítě.