

Projekt: Technologická agentura České republiky, program BETA
Číslo projektu: TITDMVCR701

Řešitel: LightComp v.o.s., Drahobejlova 1452/54, 19000 Praha

Výsledek projektu: Certifikovaná metodika

Napojení digitálních dat v prostředí Elza

Autoři: Petr Pytelka, Tomáš Pytelka (LightComp v.o.s.)
Datum: srpen 2020

1 Účel metodiky

Metodika definuje postup umožňující napojování digitálních dat na archivní popis vznikající v aplikaci Elza. Součástí postupu je vymezení požadavků na způsob práce s napojovanými digitálními daty a to v rámci jejich životního cyklu v kontextu aplikace Elza. Postup vychází z možností aplikace Elza, možností integračních rozhraní, obecných možností digitálních úložišť a požadavků na ně. Metodika přímo navazuje a je specializací dokumentu: *Metodika implementace pořadatelského software ELZA v paměťových institucích*. Současně reflektuje stávající metodiky týkající se digitálních dat v archivech, metody zpracování archiválií a cíle archivní komunity umožnit přístup k digitálním datům badatelské veřejnosti. Metodika předpokládá, že archiv má vybráno úložiště s nímž bude integrace provedena. V případě, kdy instituce implementaci úložiště připravuje, je metodika jedním ze zdrojů požadavků na funkce úložiště.

Digitálními daty pro účely metodiky se primárně rozumí digitalizáty. Předmětem metodiky není popis připojování digitálních archiválií.

1.1 Zkratky

Zkratka	Význam
API	<i>Application Programmer Interface</i> - programátorské rozhraní
DAO	<i>Digital archival object</i> - digitální archivní objekt
REST	<i>Representational state transfer</i> - metoda návrhu a způsobu interakce distribuovaných systémů
WSDL	<i>Web Services Description Language</i> - jazyk popisu rozhraní založený na XML, který se používá k popisu funkcí nabízených webovou službou
XML	<i>Extensible Markup Language</i> - značkovací jazyk pro tvorbu strukturovaných a semi-strukturovaných dokumentů

XSD	XML Schema, obvykle XML Schema 1.1
-----	------------------------------------

1.2 Definice pojmů

Terminologie použitá v této metodice vychází z terminologie definované ve standardu TOGAF a jeho oficiálního českém slovníku “TOGAF® 9.1 Translation Glossary: English – Czech” [3]. Archivní terminologie vychází z termínů definovaných v zákoně č. 499/2004 Sb. a vyhlášce č. 645/2004 Sb. Vybrané termíny jsou zopakovány v tabulce níže:

Termín	Význam
archivní soubor	Archivní fond či archivní sbírka dle 499/2004 Sb.
digitální data	V metodice použito jako synonymum k digitální archivní objekt
digitální archivní objekt	Digitální archiválie (digital born) a Digitální reprodukce archiválií (digitalizáty).
instituce / paměťová instituce	Obvykle archiv, obecně organizace zpracovávající archiválie
obecné řešení	Typový aplikační software použitý jako základ implementace, při implementaci z něj vznikne příslušný stavební blok, podrobněji viz TOGAF 2.6 Enterprise Continuum, termín <i>generic solution</i> a přechod k <i>solution building block</i>
SLA	Service Layer Agreement - smlouva o poskytování služeb. Smlouva může být uzavřena s externím subjektem nebo i v rámci organizace mezi odděleními jako neformální dohoda.
stavební blok	Připravená IT komponenta, kterou lze použít jako součást celkové architektury - viz TOGAF 3.21 Building Block a 3.66 Solution Building Block
úložiště	Technické řešení (hardware/software) umožňující dlouhodobé uložení digitálních dat
zainteresovaný subjekt	Jednotlivec, tým nebo organizace, která má zájem na výsledku uvažované změny / implementace Podrobněji viz TOGAF 3.68 Stakeholder
zájmy	Klíčové požadavky, které jsou z pohledu zainteresovaných subjektů kriticky důležité při akceptaci díla. Zájmy se mohou týkat různých aspektů: funkce, vývoj, provoz, výkonnost, spolehlivost, bezpečnost apod. Podrobněji viz TOGAF 3.30 Concerns

2 Jak s metodikou pracovat

Metodika je obecným návodem pro implementaci systému Elza ve spojitosti s úložišti digitálních dat. Tento návod je nutné adaptovat na podmínky konkrétní paměťové instituce. Jednotlivé instituce se významně liší velikostí, vnitřním uspořádáním, mají odlišné zřizovatele a odlišnou zralost procesů a jejich řízení. Adaptaci metodiky lze provést před vlastní přípravou implementace nebo jako jednu z jejích první částí.

2.1 Související metodiky

Metodika je specializací certifikované metodiky: *Metodika implementace pořadatelského software ELZA v paměťových institucích* [6]. Metodika má shodnou strukturu a je na ni možné pohlížet jako na její prohloubení v oblasti napojení digitálních dat.

Pokud je implementace prováděna v organizaci s vyšším stupněm zralosti IT procesů (dle CMMI [4]), je vhodné metodiku uzpůsobit zavedeným standardům a jim i implementaci podřídit. V opačném případě je vhodné metodiku použít jako výchozí bod pro vytvoření plánu implementace a řízení jejího průběhu.

Pokud je v instituci rozvíjena architektura dle rámce TOGAF, je metodika jedním z dokumentů popisujících plán implementace aplikace Elza jako stavebního bloku řešení (*Solution Building Block*) ve fázi *E. Příležitosti a řešení (Opportunities and Solutions)*. Metodiku je možné používat samostatně (bez rámce TOGAF). Proto důsledně nenavazuje na všechny architektonické dokumenty TOGAF a výchozí požadavky metodiky jsou minimalizovány.

2.2 Struktura metodiky

Metodika je rozdělena do několika částí:

- Fáze A - Příprava implementace
- Fáze B - Provedení implementace
- Fáze C - Provoz a další rozvoj
- Fáze D - Migrace dat

Na každou část je možné nahlížet jako na samostatnou fázi implementace. První tři části tvoří základní implementaci napojení a přímo na sebe navazují. Další fáze se již podrobněji věnují vybraným problémům a je možné je provádět nezávisle na základní implementaci.

Každá fáze je popsána strukturovanou formou. Popis začíná definicí cílů, a stručným popisem postupu, následuje popis vstupů (tj. předpokladů pro zahájení dané části implementace) a výstupů. Popis fází odpovídá způsobu popisu v TOGAF.

2.3 Užití metodiky

Metodika je návrhem postupu transformace obecného řešení na specifické pro danou instituci. Výsledkem transformace je sada stavebních bloků řešení, sada dokumentů popisujících konkrétní způsob implementace a zasazení aplikace do aplikační architektury instituce.

Implementační fáze jsou na sobě nezávislé z hlediska toho, jakou formou budou realizovány – zda vlastními silami instituce či externím dodavatelem. Metodika je neutrální i z pohledu způsobu nasazení aplikace Elza, tj. zda se jedná o hostovanou službu, provoz na vlastním hardware s externím dodavatelem, či implementaci vlastními silami.

Při přípravě implementace i v jejím průběhu by měly být samostatně uvažovány čtyři základní architektonické vrstvy, tj. organizační (business), datová, aplikační a technologická¹. Z důvodu časových, finančních a dalších omezení k tomu však často reálně v organizacích nedochází a je nutné najít vhodný vyvážený pohled. Z hlediska implementace aplikace je výchozím bodem znalost okolí uvažovaného systému v uvedených vrstvách.

2.4 Zájmy a zainteresované osoby

Při přípravě jednotlivých fází implementace je nutné zohledňovat zájmy zainteresovaných subjektů. Tyto zájmy jdou obvykle napříč všech architektonických vrstev. V rámci definice vize projektu nebo nejpozději v době jeho přípravy je nutné dotčené zájmy a jednotlivé zainteresované subjekty identifikovat.

Konkrétními příklady zájmů ve vztahu k Elza a úložišti digitálních dat:

- ekonomika / finanční hledisko
- legislativa a metodika (zahrnuje i různá vnitřní pravidla instituce)
- lidské zdroje
- bezpečnost (data, přístup k nim, provádění změn)
- ICT infrastruktura a požadavky IT oddělení
- publikace archivních pomůcek a digitalizátů

Konkrétní příklady zainteresovaných subjektů:

- archivář zpracovávající archiválie
- archivář spravující archivní soubor
- vedoucí oddělení/archivu zodpovědný za pořádání a tvorbu archivních pomůcek
- metodik archivu
- ředitel instituce
- správce publikace
- Odbor archivní správy a spisové služby MVČR jako orgán dohledu
- pracovník IT

3 Fáze A - Příprava implementace

3.1 Cíle

Cíle přípravné fáze jsou:

- stanovení rozsahu implementace / integrace
- příprava projektu implementace napojení Elza a úložiště

3.2 Způsob dosažení cíle

Fáze se zaměřuje na vytvoření projektu napojení aplikace Elza na úložiště digitálních dat a stanovení jeho rozsahu. Výchoziskem je prvotní představa business požadavků propojení

¹Podrobněji viz TOGAF 2.3 What Kind of Architectures Does TOGAF Deal With? A 5.5 Scoping the Architecture

archivního popisu a digitalizátů. Při jejich určení se vychází z dostupné dokumentace aplikace, stávajících procesů a postupů v instituci při zpracování archiválií a jejich digitalizaci. Dalším východiskem pro přípravu projektu je současná IT architektura a vize jejího dalšího rozvoje. Předmětem zájmu je aplikační, datová a technická architektura souvisejících systémů, s nimiž bude požadována integrace nebo, které budou sloužit pro provoz výsledného řešení.

Na základě vstupních dokumentů je možné začít realizovat kroky vedoucí ke stanovení rozsahu projektu. Tím se rozumí určení nákladů projektu, časového rozsahu, určení způsobu realizace a určení náročnosti na vlastní kapacity či jiné zdroje.

3.3 Vstupy

V kapitole jsou uvedeny vstupy pro Fázi A.

3.3.1 Externí dokumenty

- Technická dokumentace Elza (zejména části týkající se DAO, definice WSDL rozhraní pro úložiště)
- Technická dokumentace úložiště digitálních dat
- Legislativa související se zpracováním a zpřístupňováním archiválií

3.3.2 Interní dokumenty

- Vize / strategie instituce, zejména vize digitalizace archiválií
- Interní metodické pokyny pro zpracování archiválií
- Interní metodické pokyny pro provádění digitalizace
- Interní metodické pokyny pro uchování digitalizátů

3.3.3 Architektonické dokumenty

- Organizační struktura - s uvedením zpracovatelů archiválií, osob zodpovědných za proces digitalizace a dalších souvisejících rolí
 - Součástí by měl být i popis stávajících dotčených procesů
- Současná IT architektura - včetně popisu souvisejících systémů (aplikační, datová, technická)
 - Informace o způsobu nasazení Elza
 - Informace o způsobu nasazení Úložiště digitálních dat
- Vize budoucí IT architektury - případně základní koncept vize architektury
- Koncept specifikace požadavků
- Konceptuální představa integrace

3.4 Postup

3.4.1 Uzpůsobení metodiky

Metodika je nástrojem řízení implementace a definuje její základní rámec. Metodiku je možné upravit tak, aby odpovídala standardům používaným v organizaci. Typickou úpravou metodiky

je obsahové upřesnění formalizovaných výstupů i s ohledem na možnost využití již existujících dokumentů.

V případě implementace ve velmi malé instituci je vhodné metodiku zjednodušit. Příkladem takového zjednodušení je náhrada jednotlivých formalizovaných výstupů jejich kontrolním soupisem s poznámkami.

3.4.2 Návrh změn a úprav procesů

V organizaci by měly být definovány procesy, které pokrývají činnosti souvisejících s provozem aplikace a úložiště. V souvislosti se zavedením integrace musí dojít k jejich ověření a případné úpravě. Typické činnosti / procesy, které musí být zajištěny a jsou významné pro napojení:

- Správa archivních souborů
- Zpracování a pořádání archiválií
- Zveřejňování informací veřejnosti a badatelům
- Proces digitalizace
- Řešení dlouhodobého uchování dat a zejména digitalizátů
- Činnosti související s technickou správou aplikací - viz fáze C

Případné provedené procesní úpravy jsou zachyceny v dokumentu "Návrh změn a úprav procesů" a odpovídajících architektonických schématech.

3.4.3 Plán implementace napojení

Tvorba plánu implementace je ústředním krokem fáze A. Při jeho přípravě se vychází z konceptu požadavků. V něm jsou zohledněny zájmy jednotlivých zainteresovaných subjektů. V průběhu tvorby plánu je možné požadavky dále doplnit a upřesnit s ohledem na možnosti Elza, integračního API a možnosti úložiště. Při tvorbě seznamu požadavků je nutné zvážit využití přímo připravených funkcí a postupů v typové instalaci aplikací oproti nákladům a přínosům úprav na míru.

Při tvorbě plánu by měly být diskutovány mimo jiné tyto oblasti specializace:

- data v úložišti
 - rozsah uložených dat v úložišti, určení která budou napojena na jednotky popisu v Elza
 - zejména je nutné uvažovat:
 - vazba digitálního objektu na archivní soubor
 - rozsah popisných metadat, požadavky na jejich přenos
 - rozsah technických metadat, požadavky na jejich přenos
 - formáty souborů v úložišti a možnosti nahlížení na ně
 - kvality uložených digitalizátů a metoda práce s nimi
- technický přístup k úložišti
 - API poskytovaná úložištěm
 - možnost rozšíření úložiště o komunikační rozhraní pro Elza
 - možnost synchronizace s Elza při změně dat (směr a rozsah výměny dat)
- způsob mapování dat do Elza
 - způsob převodu podoby uložení dat v úložišti na podobu balíčků a digitálních objektů dle definice schématu WSDL, konkrétně entity Dao, DaoPackage
 - přístup k digitálním objektům z Elza, tj. přístup k celému objektu, přístup k jednotlivým stránkám

- dostupnost náhledů
- uživatelé a uživatelské role
 - možnosti jednotné autentizace do Elza a úložiště
 - způsobu řízení uživatelských účtů
 - organizace uživatelů do skupin
 - způsob provádění změn rolí v kontextu obou systémů

Na základě výše uvedených zjištění je připraven vlastní plán implementace, tj. specifikace činností v čase, odhad jejich rozsahu (pracnosti), určení, jakou formou a kdo je bude realizovat (interně či dodavatelsky). Zpracovaný plán implementace umožňuje stanovit finanční náklady. Stanovení se provede na základě znalosti jednotkových cen a předpokládaného rozsahu jednotlivých činností.

3.4.4 Zpřístupňování aplikace uživatelům

Při plánování implementace jsou v řídicím výboru projektu také zastoupeni budoucí uživatelé. V tomto kontextu je vhodné určit jednoho nebo několik klíčových uživatelů. Jedním z těchto uživatelů je obvykle pracovník/metodik zajišťující soulad způsobu zpracování archiválií s obecně závaznými oborovými standardy a metodickými pokyny instituce. Napojení by měli začít užívat jako první klíčoví uživatelé. Ti jsou schopni dále v organizaci předávat a udržovat znalosti týkající se aplikace a jejího nasazení.

Do plánu implementace je nutné zahrnout plán postupu zpřístupňování uživatelům, určení klíčových uživatelů, způsobu školení uživatelů a koncepci plánu pořádání v souvislosti s plánovanou migrací dat.

3.5 Výstupy

Výstupy pro fázi A mohou být (nikoliv výhradně):

- Uzpůsobení metodiky potřebám instituce
- Požadavky a cíle integrace
- Návrh změn a úprav procesů
 - vzniká jen pokud dochází ke změnám
- Technická zadávací dokumentace pro fázi provedení implementace
 - seznam požadavků a změn v aplikacích
 - případy užití, scénáře interakce mezi Elza a úložištěm
 - použité metody autentizace a autorizace
- Plán implementace napojení a migrace dat (časový a finanční plán)
- Upravená vize architektury IT
 - začlenění nových stavebních bloků
- Koncept plánu připojování uživatelů k aplikaci
 - zahrnuje plán způsobu zaškolení uživatelů

Technická zadávací dokumentace musí být velmi specifická. Určitě by měla detailněji popsat a definovat:

- která integrační API Elza budou užita
- rozsah metadat na úrovni digitalizátu (v úložišti, předávaná do Elza)
Příklad: textový popis obsahu digitalizátu, datace obsahu

- rozsah technických metadat digitálních objektů (v úložišti, předávaná do Elza)
Příklad: rozlišení digitalizátu
- způsob zobrazení digitálních objektů
Je nutné určit jakým způsobem si může archivář digitální objekt zobrazit, jaký způsobem je tento zobrazen v Elza
- metody řešení synchronizace při změnách
- způsob předávání identifikátorů digitálních objektů ve výstupech

4 Fáze B - Provedení implementace

4.1 Cíle

Cíle fáze provedení implementace jsou:

- vlastní realizace připraveného implementačního projektu a vytvoření stavebních bloků řešení
- převzetí implementace a akceptace do produkčního prostředí
- zahájení využívání připraveného řešení
- vyškolení prvních uživatelů

4.2 Způsob dosažení cíle

Na základě připraveného projektu je spuštěn projekt implementace. Pro řízení projektu se doporučuje použít vhodnou metodiku v souladu s pravidly instituce, kde implementace probíhá (například PMBOK [5], PRINCE2). Formou kontrolních dní je sledován průběh projektu a jsou také řešeny případné požadavky na změnu. Správné provedení implementace je ověřeno formou akceptačního protokolu.

4.3 Vstupy

V kapitole jsou uvedeny vstupy pro Fázi B.

4.3.1 Externí dokumenty

- Technická dokumentace Elza (zejména části týkající se DAO, definice WSDL rozhraní pro úložiště)
- Technická dokumentace úložiště digitálních dat

4.3.2 Interní dokumenty

- Plán implementace napojení dle Fáze A (časový a finanční plán)
- Koncept plánu zpřístupnění napojení uživatelům (v Elza i úložišti)

4.3.3 Architektonické dokumenty

- Uzpůsobení metodiky potřebám instituce
- Návrh změn a úprav procesů

- Technická zadávací dokumentace pro fázi provedení implementace
- Vize architektury IT

4.4 Postup

4.4.1 Zahájení implementace

Prvním krokem fáze provedení implementace je stanovení metodiky řízení projektu. Iniciačním krokem je vytvoření řídicího výboru projektu, kde budou jednotliví participanti (sponzor, zástupce uživatelů, řešitel projektu - ať již interní či externí). Při zahájení projektu je nutné stanovit časový harmonogram řešení, způsob a pravidla komunikace v průběhu řešení, kontrolní dny, způsob předání řešení a pravidla akceptace řešení. V případě externího dodavatele jsou uvedená pravidla součástí smlouvy o dílo.

4.4.2 Sledování průběhu implementace a akceptace řešení

Formou kontrolních dní je obvykle sledován průběh projektu. Aktuální zjištění a změny v rozsahu projektu musejí být formalizovány (například formou požadavků na změnu - *change request*) dle použité metodiky řízení.

Na základě definovaných požadavků je zpracován akceptační protokol. V průběhu předávání je nutné verifikovat funkce napojení a splnění jednotlivých požadavků zadání. Vhodnou formou pro ověření všech funkcí je "pilotní provoz". Po určitou definovanou dobu je možné napojení používat v režimu se zvýšeným dohledem, obvykle pro menší počet uživatelů pro finální ověření všech funkcí před plošným nasazením. Pro ověření v průběhu pilotního provozu je nutné zvolit vhodný archivní soubor.

Již v průběhu implementace může vznikat seznam budoucích změn a rozšíření. Při implementaci je vhodné se soustředit na zprovoznění napojení v produkčním prostředí. Nové nekritické požadavky, které při implementaci vzniknou, je možno formou iterativního vývoje realizovat následně.

4.5 Výstupy

Výstupy pro fázi B mohou být (nikoliv výhradně):

- Smlouva o provedení implementace s externím dodavatelem
- Akceptační protokol
- Realizované napojení v testovacím prostředí
- Realizované napojení v produkčním prostředí
- Seznam budoucích změn a rozšíření

5 Fáze C - Provoz a další rozvoj

5.1 Cíle

Cíle fáze provozu a dalšího rozvoje jsou:

- zajištění metodické, odborné a technické podpory uživatelů systému

- zajištění změn způsobu napojení digitálních dat do systému Elza v závislosti na měnících se podmínkách a požadavcích
- zajištění způsobu řešení a oprav chybových stavů
- zajištění odborného zaškolování nových uživatelů

5.2 Způsob dosažení cíle

Provozní parametry definuje IT oddělení instituce v kooperaci s vlastníky dotčených procesů. V rámci dlouhodobé udržitelnosti je nutné pravidelně vyhodnocovat reálné parametry služby, požadavky na změny a přijímat příslušná opatření. Na úrovni napojení digitálních dat je zejména nutné sledovat stav napojení a synchronizace informací mezi systémy.

5.3 Vstupy

V kapitole jsou uvedeny vstupy pro Fázi C.

5.3.1 Interní dokumenty

- Uživatelé v Elza a úložišti digitálních dat

5.3.2 Architektonické dokumenty

- Současná a budoucí architektura IT
- Seznam budoucích změn a rozšíření

5.4 Postup

5.4.1 Zajištění funkčního napojení Elza a úložiště

Předpokladem je zajištění fungování obou aplikací v rámci jejich SLA. Pro zajištění funkčního napojení je vhodné definovat metodu jeho diagnostiky, nejlépe automatizovanou kontrolu. Průběžně je nutné ověřovat stav synchronizace dat a to při rutinní práci uživatelů a při mimořádných zásazích a změnách. V případě zjištění nesouladu se provedou nápravná opatření ve dvou rovinách:

- obnovení požadovaného stavu synchronizace dat
- identifikace příčiny vzniku chybového stavu a přijmutí nápravných opatření (například formou požadavku na změnu)

5.4.2 Poskytování uživatelské a technické podpory

Podpora je poskytována obvykle samostatně pro aplikaci Elza i úložiště. Na úrovni napojení obou systémů je důležité zajistit přenos informací mezi osobami zodpovědnými za jednotlivé systémy a jejich vzájemnou součinnost při řešení chyb.

5.4.3 Řízení změn, aktualizací aplikací a napojení

Elza i úložiště jsou obvykle aktualizovány a měněny nezávisle na sobě. Při plánování změn a aktualizací je nutné ověřovat jejich dopad na napojení. Dopad je nutné ověřovat v rovině metodické i technické.

Požadavky na změnu napojení je nutné řešit formalizovaným požadavkem. Způsob jeho zápisu a vyřízení je obvykle definován interními procesy IT oddělení a jím používaných metodik řízení.

Je vhodné, aby v instituci byly k dispozici minimálně dvě prostředí aplikací. Jedno s produkčními daty a druhá testovací. Testovací verze slouží k ověření nových funkcí, způsobu užití, školení uživatelů apod. Nové verze by vždy měly být nejprve instalovány do testovacího prostředí a až následně do produkčního.

5.5 Výstupy

Výstupy pro fázi C mohou být (nikoliv výhradně):

- Metody zajištění a ověřování funkčního napojení Elza a úložiště
- Výsledky pravidelných kontrol napojení, integrity dat
- Definice způsobu hlášení incidentů/chyb a to s ohledem na spolupráci zodpovědných osob
- Upravená SLA pro Elza a digitální úložiště
- Změnové požadavky

6 Fáze D - Migrace dat

6.1 Cíle

Cíle fáze migrace dat jsou:

- provedení převodu napojení digitálních dat
- ukončení využívání staršího systému

6.2 Způsob dosažení cíle

Východiskem pro migraci napojení digitálních dat je jejich existence v úložišti. Pokud existují vně úložiště, je nutné samostatně definovat metodu jejich vložení do úložiště.

Předmětem migrace napojení je přenos identifikátorů DAO v úložišti a jejich napojení na jednotky popisu v Elza. Je nutné stanovit průběh a formu tohoto přenosu.

6.3 Vstupy

V kapitole jsou uvedeny vstupy pro Fázi D.

6.3.1 Externí dokumenty

- Dokumentace aplikace Elza (zejména definice API)
- Dokumentace úložiště (zejména definice API)

6.3.2 Interní dokumenty

- Popis stávajících pravidel pro pořádání
- Rozsah metadat a technických metadat na úrovni digitalizátu
- Datové soubory pro migraci (nativní datový formát původní aplikace)

6.3.3 Architektonické dokumenty

- Popis používaných aplikací pro zpracování archiválií, resp. datových formátů a verzí těchto formátů
- Vize IT infrastruktury

6.4 Postup

6.4.1 Příprava migrace

Při přípravě migrace dat je prvním krokem klasifikace migrovaných dat, tj. rozdělení do tříd dle předpokládaného způsobu migrace. Dělení bude například dle zdrojové aplikace, dalších požadavků na budoucí zpracování migrovaných dat, použitých pravidel pro pořádání apod. Dále je nutné zvážit, zda migrovaná data budou mít formu samostatného archivního souboru nebo tvoří část existujícího souboru. Při přípravě projektu je dále nutné zhodnotit, které prvky popisu budou použity, případně navrhnout zavedení dalších. Při přípravě migrace je nutné stanovit, které kroky a zda budou probíhat ručně, částečně automatizovaně, automatizovaně. Na základě výše uvedených podkladů je možné stanovit náklady spojené s migrací dat a také připravit časový harmonogram migrace. Současně je vhodné zvážit, kdo migraci či její část provede, tj. zda vlastními silami nebo pomocí externího dodavatele.

6.4.2 Provedení migrace

Provedení migrace by měl být již relativně mechanický postup dle připraveného plánu. Pokud migrace vyžaduje ruční zásahy, tak je nutné mít stanovené odpovědné osoby, které je provedou. Součástí provedené migrace dat by také měl být protokol o jejím provedení.

6.4.3 Kontrola migrovaných dat

Pokud dochází k trvalému převodu dat do nového formátu, tak je nutné stanovit kontrolní mechanismy pro zajištění kvality převodu. Příkladem základních kontrol jsou metriky: počet napojovaných digitalizátů, počet vytvořených napojení, počet nových či aktualizovaných jednotek popisu apod.

Dále je vhodné provést namátkovou kontrolu struktury, zda odpovídá očekávání.

6.5 Výstupy

- Plán převodu dat
- Jednotky popisu s platně napojenými digitalizáty
- Protokol o provedeném napojení

7 Novost postupů

Metodika popisuje postup napojení digitálních dat v prostředí aplikace Elza. Metodika vychází ze stávající metodiky implementace Elza a tuto významně prohlubuje v kontextu napojení na digitální úložiště. Metodika je určena pro všechny typy archivů a jejich vedoucí pracovníky odpovědné za implementaci napojení.

Metodika se uplatní při implementaci napojení digitálních dat ve všech jejích fázích, tj. příprava, vlastní technické provedení implementace, zajištění dlouhodobé udržitelnosti a migrace dat.

Přínosem použití metodiky pro uživatele je zajištění implementace napojení v plné šíři a to tak, aby byly zohledněny zájmy všech zainteresovaných subjektů. Metodika poukazuje v jednotlivých fázích na jednotlivé aspekty, které je nutné zvážit a do projektu zahrnout. Metodiku je možné používat opakovaně a tak dále rozsah implementace zvětšovat.

8 Seznam literatury

[1] WANNER, Michal. *Základní pravidla pro zpracování archiválií*. Druhé, doplněné a rozšířené vydání. Praha: Odbor archivní správy a spisové služby MV, 2015. ISBN 978-80-86466-78-1.

[2] THE OPEN GROUP ARCHITECTURE FORUM. *TOGAF Version 9.1*. Van Haren Publishing, 2011. ISBN 978908753679C. Dostupné také z: <http://pubs.opengroup.org/architecture/togaf9-doc/arch/>

[3] *CMMI for Services, Version 1.3: Improving processes for providing better services* [online]. Pittsburgh, Pennsylvania, United States: Carnegie Mellon University, 2010 [cit. 2016-10-28]. Dostupné z: http://cmmiinstitute.com/system/files/models/CMMI_for_Services_v1.3.pdf

[4] *TOGAF® 9.1 Translation Glossary: English – Czech* [online]. The Open Group, 2013 [cit. 2016-10-30]. ISBN 1-937218-33-1. Dostupné z: <http://www.mvcr.cz/soubor/oficialni-togaf-slovník-pdf.aspx>

[5] *A guide to the project management body of knowledge: (PMBOK® guide)*. Fifth edition. Newtown Square: Project Management Institute, 2013. ISBN 978-1-935589-67-9.

[6] *Metodika implementace pořadacího software ELZA v paměťových institucích* [online]. listopad 2016 [cit. 2020-05-01]. Dostupné z: <https://www.mvcr.cz/soubor/metodika-implementace-poradaciho-software-elza-v-pametovych-institucich.aspx>

9 Příloha 1 - Ukázka části technické dokumentace napojení Elza - DSpace

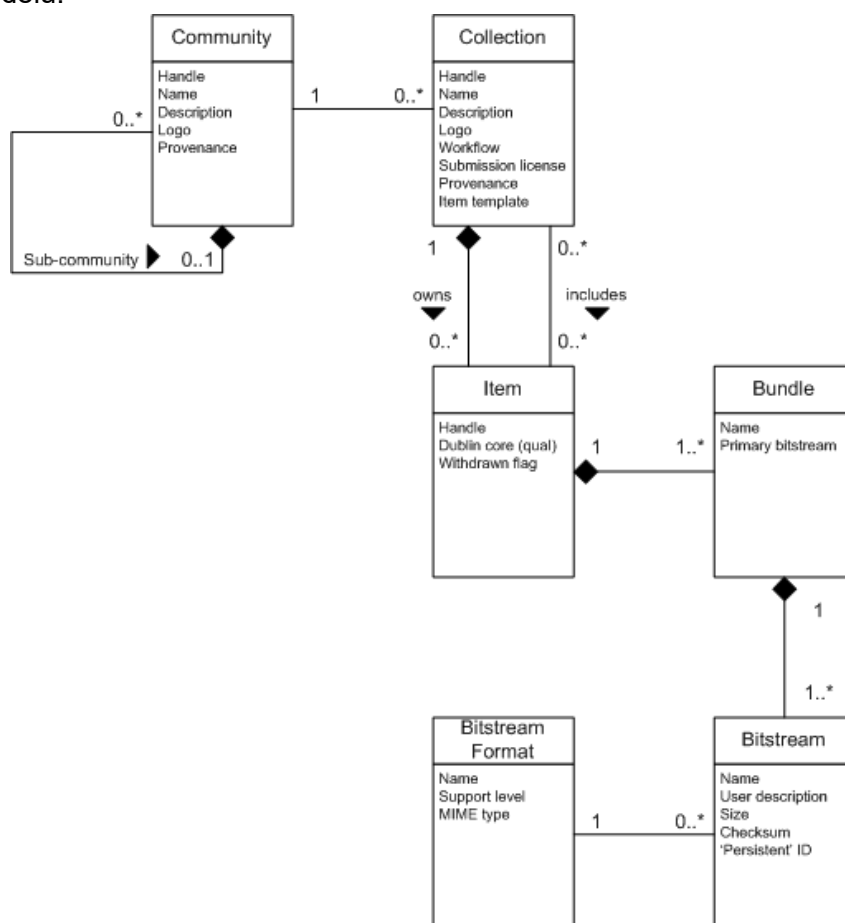
Příklad části technické dokumentace napojení aplikace Elza a úložiště DSpace.

9.1 Data v úložišti

Východiska pro provedení datové integrace jsou:

- Data (digitalizáty) v úložišti jsou organizována dle příslušnosti k archivnímu souboru
- Předmětem uložení jsou digitalizáty; digitalizátem rozumíme jeden nebo více souborů tvořících posloupnost (například: jedna fotografie, jednotlivé stránky knihy)
- Součástí uložení digitalizátů jsou popisná metadata, technická metadata a případně náhledy.

Datový model DSpace je relativně přímočarý a pevný. V základní podobě pracuje s entitami Komunita (*Community*), Kolekce (*Collection*), Položka (*Item*), Svazek (*Bundle*), Datový proud (*Bitstream*), Datový formát (*Bitstream Format*). Následující schéma zobrazuje klíčovou část datového modelu:



Popis jednotlivých entit je součástí referenční dokumentace DSpace.

9.1.1 Základní struktura dat v úložišti a její užití

Komunita (*Community*) - uložení struktury archivů

Kolekce (*Collection*) - uložení archivních souborů

Položka (*Item*) - uložení jednotlivých digitalizátů v rámci kolekce

Svazek (*Bundle*) - předdefinované typy svazků (ORIGINAL - vlastní data, TECHMD - technická metadata, THUMBNAIL - náhledy)

Datový proud (Bitstream) - binární stream dat pro danou stránku ve svazku

Příklad:

```
MZA Brno
- SOkA Vyškov
  - Národní výbor Vyškov
- SOkA Hodonín
```

Poznámka: Je možné vytvářet i virtuální kolekce využívající data uložená v jiných kolekcích. Data jsou uložena jen 1x. Může být využito například pro tematické katalogy či jiné soupisy. V datovém úložišti je možné ukládat i jiná data než-li jsou digitalizáty příslušné k archivním souborům. Příkladem mohou být data pro knihovnu archivu, jiné skeny apod. Pro uložení těchto dat lze využít dvou mechanismů: založení vlastní komunity (komunit) nebo vytvoření jiné kolekce na úrovni archivu. V obou případech entita nesmí svými případnými identifikátory být zaměněna s archivem či archivním souborem. Žádná jiná omezení na ní nejsou kladena.

9.1.2 Komunita a její metadata

Komunity se používají pro popis archivů. Komunity jsou hierarchické a touto formou lze reprezentovat strukturu archivů v případě státních oblastních archivů. Pro komunitu je bez úprav k dispozici jen omezená množina metadat:

- Název (name) - ukládá se jako DC:title
- Copyright text (copyright text) - ukládá se jako DC:rights
- Úvod (introduction text) - ukládá se jako DC:description
- Stručná charakteristika (short description) - ukládá se jako DC:description:abstract
- Text pro boční panel (sidebar text) - ukládá se jako DC:description:tableofcontents
- Logo

Archiv je v úložišti reprezentován pomocí metadat:

- **Název archivu** - uloží se do pole *Název (name)*
- **Číslo archivu** - uloží se do pole *Stručná charakteristika (short description)*

9.1.3 Kolekce a její metadata

Kolekce slouží pro popis archivního souboru. Kolekce je součástí právě jedné komunity, tj. Archivu. Pro popis kolekce je bez úprav k dispozici jen omezená množina metadat:

- Název (name) - ukládá se jako DC:title
- Copyright text (copyright text) - ukládá se jako DC:rights
- Úvod (introduction text) - ukládá se jako DC:description
- Stručná charakteristika (short description) - ukládá se jako DC:description:abstract
- Text pro boční panel (sidebar text) - ukládá se jako DC:description:tableofcontents
- Logo

Uživatelské rozhraní DSpace/XMLUI navíc umožňuje vkládat pole *Původ (provenance)*. Toto pole není dostupné pomocí REST API.

Archivní soubor je v úložišti reprezentován pomocí metadat:

- **Název archivního souboru** - uloží se do pole *Název (name)*
- **Číslo AS / NAD** - uloží se do pole *Stručná charakteristika (short description)*

Další pole je možné volitelně vyplnit dle uvážení archivu, nejsou však předmětem integrace s Elza.

9.1.4 Digitalizáty a jejich metadata

Do DSpace se ukládají jednotlivé digitalizáty. Digitalizátem rozumíme jeden nebo více souborů, které tvoří digitální obraz archiválie nebo její části - výsledek digitalizace. Digitalizát je zařazen do právě jedné kolekce (archivního souboru) a je popsán sadou metadat. Digitalizát je tvořen jednotlivými datovými soubory, technickými metadaty a případně připravenými náhledy.

Rozsah metadat na úrovni položky/digitalizátu:

- Název - textový popis položky, obvykle bude odpovídat regestu v archivním popisu
- Identifikátor digitalizátu v DSpace
- Identifikátor archivního popisu pokud je k němu digitalizát připojen (UUID)
- Referenční označení archivního popisu k němuž je digitalizát připojen (pokud existuje)
- Datace vzniku
- Datace pořízení digitalizátu
- Identifikace vstupní dávky

Za účelem uložení binárních dat souborů, technických metadat a náhledů jsou definovány tyto typy svazků:

- ORIGINAL - primární data tvořící digitalizát
- METADATA - doplňující metadata k primárním datům, obvykle technická metadata stránek
- THUMBNAIL - náhledy na data tvořící digitalizát

Uvedené typy svazků jsou obvykle využívány v DSpace a je pro ně případně k dispozici základní podpora v uživatelském rozhraní.

Svazek ORIGINAL slouží pro uložení vlastních stránek digitálního dokumentu. Součástí uložení jsou tato metadata:

- Název souboru / stránky (*name*)
- Popis souboru (*description*)
- Mime-type stránky (*mimetype*)
- Checksum
- Pořadí souboru/stránky
- Velikost v bytech

Formát dat ukládaných do svazku ORIGINAL by měl být dán vnitřní metodikou archivu. V případě rastrových obrázků jsou příkladem vhodného formátu pro uložení TIFF, JPEG2000, WEBP apod.

Svazek METADATA slouží pro uložení metadat jednotlivých stránek digitálního dokumentu. Pro párování souborů mezi svazky se použije jejich název (*name*). Metadata se v závislosti

na datovém formátu ukládají v XML formátu. Typ XML je uložen v poli s popisem souboru (*description*). Příпустné formáty metadat jsou:

- Pro obrázky: MIX Schema Version 2.0 (<http://www.loc.gov/standards/mix/>), do description se uvádí: MIX
- Pro audio: AudioMD version 2.0 (<https://www.loc.gov/standards/amdvmd/index.html>), do description se uvádí audioMD
- Pro video: VideoMD version 2.0 (<https://www.loc.gov/standards/amdvmd/index.html>), do description se uvádí videoMD

Svazek THUMBNAIL slouží pro uložení náhledů na originální soubory. Náhledy se ukládají ve formátu JPEG. Jednotlivé náhledy jsou párovány k originálům pomocí svého názvu.

9.2 Technický přístup k úložišti

Přístup k úložišti je zajištěn pomocí integrační komponenty. Ta komunikuje s Elzou pomocí standardního WSDL rozhraní. Integrační komponenta je samostatná webová aplikace založená na projektu DSpace. Komponenta implementuje WSDL služby Elza pro práci s digitalizáty a současně umožňuje vkládat a vyhledávat data v systému DSpace. Integrační komponenta disponuje uživatelským rozhraním umožňujícím realizovat jednotlivé operace.

9.3 Způsob napojení dat do Elza

Případy užití a způsob řešení jejich podpory:

Užití	Způsob řešení
Přijetí požadavku na digitalizaci	Není podporováno
Vkládání digitalizátů	Vstupní adresář odkud jsou čteny nové digitalizáty a jejich metadata
Zobrazení digitalizátu	Digitalizáty se zobrazují pomocí permanentních odkazů do xmlui (příklad: http://localhost:8070/xmlui/handle/123456789/5). Jednotlivé soubory je možné také zobrazit.
Zaslání nepřípojených digitalizátů	V UI úložiště je možné do Elza zaslat jeden konkrétní digitalizát pro připojení nebo pomocí dotazu vybrat celou skupinu a tu zaslat.
Uložení informace o připojení/odpojení digitalizátu	Přijetí informace o připojení/odpojení digitalizátu a její uložení
Aktualizace metadat	U digitalizátů je možno aktualizovat metadata na základě hodnot nastavených v aplikaci Elza
Žádost o delimitaci	Provede se přesun mezi archivními soubory. Vstupem je úplný identifikátor AS.
Vnitřní skartace	Vymazání digitalizátu z úložiště
Zasílání notifikací z úložiště	Není podporováno, je nahrazeno možností explicitní

9.3.1 Vkládání digitalizátů

Vkládání digitalizátů bude realizováno vložením souborů v definované podobě do vstupního adresáře. Podoba souborů bude ve variantách:

- Minimální: adresář určující AS k němuž data patří, obsahující 1..n datových souborů (například JPEG) s koncovým souborem definujícím konec vstupu
- Pokročilá: navíc může být součástí XML definující metadata digitalizátu, pořadí souborů, odkaz na soubory s technickými metadaty

Pokud vkládané soubory nebudou obsahovat explicitně technická metadata (v samostatném souboru), tak budou metadata u známých formátů extrahována pomocí nástroje JHOVE (<https://github.com/openpreserve/jhove>). Metadata budou uložena do svazku METADATA dle definice výše.

Pomocí nastavení úložiště je možné vynutit kontrolu přípustnosti formátu vkládaných dat do svazku ORIGINAL. Vkládaná data musí být známého typu. Pro ověření validnosti vkládaných dat lze využít nástroj DROID (<https://digital-preservation.github.io/droid/>). Příkladem užití je požadavek na jednotný ukládací formát pro obrázky (například JPEG 2000). Volitelnou funkcí je možnost automatizované konverze do ukládacího formátu s užitím externí aplikace (například automatizace převodu TIFF na JPEG 2000).

Součástí vložení je automatické generování náhledů do svazku THUMBNAIL dle konfigurace. V rámci integrační komponenty musí být zajištěno, že po vložení souborů dojde k automatizovanému vygenerování technických metadat a náhledů pokud nejsou přímo součástí vkládaného digitalizátu. Pro zajištění těchto funkcí se předpokládá využití v maximální míře vestavěných možností DSpace, jejich konfigurace a napojení (spouštění) na základě vložení dat, popis nastavení. Teprve v případě zjištění nedostatečnosti vestavěných funkcí budou tyto naprogramovány v rámci modulu.

9.3.2 Zobrazení digitalizátu

Digitalizát je možné zobrazit jako celek, je možné zobrazit konkrétní stránku nebo náhled konkrétní stránky. Formát uložení podkladových souborů závisí na vnitřní metodice archivu a není součástí návrhu způsobu integrace. Zobrazení celého digitalizátu bude obsahovat jeho metadata včetně linku do aplikace Elza (pokud je digitalizát připojen k archivnímu popisu).

V rámci implementace je nutné zajistit zobrazení náhledu stránky na základě znalosti ID souboru/stránky. Náhledy budou standardně uloženy ve formátu JPEG. Náhledem stránky se rozumí hypertextový odkaz na obrázek, který je možné z úložiště přímo načíst a použít v rámci HTML stránky.

9.3.3 Uložení informace o připojení/odpojení digitalizátu

Při připojení digitalizátu k jednotce popisu musí být do úložiště zaslána notifikace o této akci. WSDL obsahuje příslušné metody a v Elze jsou implementovány. Součástí zasílaných informací bude UUID jednotky popisu, identifikátor digitalizátu, regist, datace, referenční označení a případná další popisná metadata.

Komponenta provede uložení (namapování) zaslaných popisných metadat do úložiště. Způsob mapování prvků popisu z Elza je definován v části [4.5 Digitalizáty a jejich metadata](#). V případě odpojení digitalizátu je o tomto zaslána notifikace do DSpace.

9.3.4 Zaslání nepřípojených digitalizátů

Digitalizát, který není připojen k jednotce popisu je možné zaslat do Elzy k napojení.

Na nepřípojené digitalizáty je možné se dotázat pomocí:

- Názvu digitalizátu
- Datace - časový rozsah
- Identifikace vstupní dávky

Výsledek dotazu je možné zaslat do Elzy. Lze limitovat počet zaslaných digitalizátů, případně vybrat odesílané nebo odeslat všechny. Digitalizáty se odesílají pomocí balíčků, kdy každý digitalizát je předáván formou samostatného balíčku.

9.3.5 Aktualizace metadat

Jednotku popisu k níž je připojen digitalizát je možné v aplikaci Elza měnit a aktualizovat. Metadata o této jednotce popisu se do úložiště přenášejí v okamžiku připojení digitalizátu. V důsledku později provedených změn v Elze však nemusejí být aktuální. Funkce pro aktualizaci metadat slouží k tomu, aby metadata v úložišti byla synchronizována s hodnotami prvků popisu v Elze.

Synchronizaci metadat musí být možné spustit ručně ze stránky připojeného digitalizátu v DSpace. Druhou možností je hromadná akce pro provedení aktualizace metadat všech připojených digitalizátů dané kolekce.

9.3.6 Delimitace

Delimitací se rozumí přesun digitalizátu mezi jednotlivými archivními soubory. Pokyn k tomuto přesunu je možné zadat z aplikace Elza pomocí webových služeb. Součástí pokynu je informace o cílovém AS. Při přesunu dochází k odstranění informace o napojení k stávajícím jednotkám popisu.

Identifikátor cílového archivního souboru se zadává jako Archiv/číslo NAD

9.3.7 Vnitřní skartace

Operace vnitřní skartace slouží k trvalému odstranění nevyhovujícího digitalizátu. Operace by měla být realizována jako vymazání z DSpace.

9.4 Uživatelé a uživatelské role

Systémy Elza i DSpace mají vlastní správu uživatelů a oprávnění k uloženým datům. V rámci integrace je nutné řešit jakým způsobem provádět koordinaci uživatelských účtů, jejich hesel. Hlavním případem užití je možnost přecházet mezi archivním popisem a navázaným digitalizátem (tento zobrazovat), kdy každá z těchto entit je uložena v jiném systému.

Možné varianty:

- Bez integrace (nezávislí uživatelé v DSpace a Elza), případně varianta, kdy entity v DSpace jsou v režimu čtení dostupné bez omezení
- Primárně uživatelé v Elza, autentizace se přebírá do DSpace
- Primárně uživatelé v LDAP - autentizace se přebírá do Elza i DSpace
- Úplně řešení s využitím nezávislého Identity Provider (např. SAM s využitím Shibboleth, CAS, ...) - autentizace se přebírá do Elza i DSpace, plná podpora SingleSignOn (tj. jedno přihlášení pro více systémů)

Z hlediska vývoje řešení a možnosti jeho reálného nasazení lze realizovat v prvním kroku integraci bez integrace uživatelských oprávnění. Z důvodu integrace obou systémů na úrovni uživatelského rozhraní bude vhodné zavést jednotnou správu uživatelských účtů v obou systémech a umožnit podporu Single Sign On (SSO). V druhém kroku realizována integrace s využitím samostatného Identity Provider (konkrétní typ dle možností organizace, např. Windows Active Directory + Federation Services).

Volání WSDL služeb v Elza bude autorizováno pomocí technického účtu. Avšak vyhodnocení dané operace proběhne na základě oprávnění uživatele, který operaci navrhuje.