

Univerzita Karlova

Filozofická fakulta

Ústav Informačních Studií a Knihovnictví

Bakalářská práce

Jan Dobiášovský

**Správa digitální knihovny Národní technické knihovny a dlouhodobá ochrana
digitálních dokumentů: případová studie**

Managing digital library of the National Library of Technology and the long-term
preservation of digital documents: a case study

Praha 2017

Vedoucí práce: PhDr. Radka Římanová, PhD.

Poděkování

Rád bych poděkoval svojí vedoucí práce PhDr. Radce Římanové, PhD za odbornou konzultaci, trpělivost a nové poznatky k mému budoucímu akademickému i profesnímu rozvoji, kolegům z Národní Technické knihovny za hodiny tvrdé, ale důležité a užitečné práce a v neposlední řadě a blízkým za obrovskou podporu a povzbuzování, kterého se mi v průběhu psaní práce dostalo.

Prohlášení:

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

V Praze, dne 28. dubna 2017

.....

Jméno a příjmení

Abstrakt (česky)

V úvodní části je obecně a stručně popsán stav standardizace v oblasti dlouhodobé ochrany digitálních dokumentů na mezinárodní úrovni. Dále je na základě dostupné literatury popsána situace v České republice, s ohledem zejména na oblast digitálního knihovnictví.

V praktické části je popsána praxe správy digitální knihovny Národní technické knihovny, konkrétně na sbírky digitálních objektů, ke kterým má NTK trvalou autorskoprávní licenci. Metodou praktické části byla analýza dokumentace vnitřních metodik a pracovních postupů, pozorování procesů informačního managementu digitální knihovny v NTK, s důrazem na preservační procesy, a polostrukturované rozhovory s odpovědnými osobami NTK. V závěrečné části je tvořena komparativní studii vybraného standardu LTP (ISO 16363) s praxí NTK s vytyčením hlavních problémových okruhů.

Klíčová slova (česky)

dlouhodobá archivace, digitální informace, digitální kurátorství, standardizace, digitální knihovny

Abstract (English):

Theoretical part of this document briefly describes current state of LTP standardization on global scale and corresponding situation in Czech Republic according to available literature with focus on digital curation.

Practical part analyses management of NTK digital library, its documentation, processes and outputs in archiving of digital documents based on half-structured interviews with responsible institution employees. As final product there will be comparative study of LTP standard of choice (ISO 16363) and practice within institution which will lay out most problematic areas.

Keywords (English):

long-term preservation, digital information, digital curation, standardization, digital libraries

1	Úvod.....	1
1.1	Mezinárodní standardy dlouhodobé ochrany digitálních dokumentů.....	1
1.1.1	Task Force on Archiving of Digital Information	2
1.1.2	Referenční rámec OAIS	3
1.2	Standardy v České republice.....	14
1.2.1	Dlouhodobá ochrana digitálních dokumentů	15
2	Audit digitální knihovny Národní technické knihovny.....	22
2.1	Metodika auditu	22
2.2	Cíle auditu.....	22
2.3	Aktuální stav digitální knihovny Kramerius NTK.....	23
2.4	Audit ISO podle ISO 16363.....	26
3	Vyhodnocení auditu digitální knihovny NTK.....	41
3.1	Výstupy auditu digitální knihovny digitální knihovny NTK.....	41
3.1.1	Krátkodobé cíle	41
3.1.2	Dlouhodobé cíle	42
4	Závěr.....	43
5	Seznam použitých zdrojů:	45
6	Seznam obrázků:	49
7	Seznam grafů.....	49
8	Seznam zkratk:	50
9	Seznam příloh.....	51

Předmluva

Digitální kurátorství je poměrně nová disciplína, která mi přišla vždy jako velice zajímavá právě tím, že sleduje a odhaluje průvodní jevy informatizace a propojuje mé osobě velmi blízké obory informačních technologií, informační vědy a knihovnictví na poli uchovávání dokumentů. Dále je tento obor důležitý, protože zajišťuje, aby digitální materiály byly uchovány a použitelné i pro budoucí generace a výzkum, protože každý dobrý výzkum potřebuje i dobrá data. Svoji bakalářskou práci bych rád přispěl k tématu digitálních knihoven a zároveň se pokusil být užitečný digitální knihovně Národní technické knihovny, kde jsem aktuálně zaměstnán jako administrátor.

NTK používá digitální knihovnu Kramerius, aktuálně ve verzi 5.2 pro veřejnost a testovacího klienta ve verzi 5.3.6. Jedná se softwarové řešení digitální knihovny pro zpřístupňování jak klasických digitalizátů tak i digital-born dokumentů průběžně upravované tak, aby struktura metadat odpovídala standardům vyhlášených Národní knihovnou České republiky. Obsahuje rozhraní umožňující vyhledávání jak v metadatech, tak plném textu, tvorbu a správu virtuálních sbírek, popř. generování vícestránkových PDF dokumentů pro volně dostupné dokumenty a administrátorské rozhraní (Kramerius, 2017). Aktuálně Kramerius obsahuje přes 1700 monografií a několik periodik z historického fondu knihovny.

1 Úvod

V současné době stojí informační instituce před řešením nového úkolu zabezpečení dlouhodobé archivace digitálních objektů. Přestože mají informační instituce k dispozici stále levnější digitální úložiště, které unesou větší objem digitálních informací, nelze archivovat kompletně. Vystavujeme se tedy nebezpečí, že nezvládneme archivovat důležité dokumenty, které by mohly mít historický význam nebo bychom nemuseli být schopni zajistit jejich vyhledávání vzhledem k inkonzistenci či absenci metadat. Velký objem dat také vyžaduje specializované nástroje pro jejich správu a zpracování a automatizaci. Dalším problémem je, že nelze uchovat některé materiály vzhledem k jejich špatné kondici, konfliktům autorského práva či povaze jejich obsahu (dokumenty obsahující osobní či důvěrné informace, bezpečnostní rizika či posvátné nebo kulturně nevhodné materiály) (Driscoll, 2006). Dalším formou rizika mohou být například duplicity digitalizovaných dokumentů. Pro udržitelnou archivaci digitálních dokumentů je také důležitá zpětná použitelnost hardware a software. V současném tempu vývoje nových technologií je potřeba držet krok jednak s trendy, ale zároveň i zajistit interoperabilitu. Zároveň i software jako prostředek zobrazení digitálního materiálu se může časem objevit v pozici kdy je potřeba jej archivovat. Digitální knihovny by zároveň měly být schopny kromě materiálu uchovat i původní uživatelský zážitek (Rieger, 2015).

Cílem této práce je popsat základní standardizující dokumenty na mezinárodní úrovni a na základě dostupné literatury pospat stav standardizace v oblasti dlouhodobé archivace dat v České republice. Na základě těchto dokumentů bude skrze vybraný nástroj analyzována správa digitální knihovny budované Národní technickou knihovnou (NTK) z hlediska standardů a strategií dlouhodobé ochrany digitálních dokumentů.

1.1 *Mezinárodní standardy dlouhodobé ochrany digitálních dokumentů*

Za dva nejvýznamnější projekty pro dlouhodobou archivaci digitálních dat jsou považovány: Task Force on Archiving Digital of Information a referenční rámec pro Open Archival Information System (OAIS) – dokument na kterém je založena norma OAIS ISO 14721:2012, na níž jsou poté založeny další příbuzné normativní dokumenty.

1.1.1 Task Force on Archiving of Digital Information

V prosinci 1994 vznikla z iniciativy *Commission on Preservation and Access and the Research Libraries Group* skupina *Task Force on Archiving of Digital Information*. Tato skupina strávila rok studiemí, jak stabilizovat digitální knihovny, učinit je více přístupné a odůvodnit jejich funkci.

1.1.1.1 Výstupy

Výsledkem výzkumu Task Force vznikla publikace *Preserving digital information: report of the Task Force on Archiving of Digital Information*. Jejími závěry byly následující doporučení:

- 1) *První linie obrany proti ztrátě důležitých digitálních informací záleží na tvůrcích, poskytovatelích a majitelích digitální informace.*
- 2) *Dlouhodobá ochrana digitální informace na adekvátní úrovni pro budoucí výzkum a vědeckou erudici bude vyžadovat propracovanou infrastrukturu schopnou provozovat distribuovaný systém digitálního archivu*
- 3) *Kritickou součástí struktury digitální archivace bude existence dostatečného množství důvěryhodných organizací schopných ukládání, migrace a poskytování přístupu k digitálním sbírkám*
- 4) *Proces certifikace digitálních archivů bude vyžadován pro vytvoření celkového důvěryhodného klima v oboru dlouhodobé ochrany digitálních dokumentů*
- 5) *Certifikované digitální archivy musí mít právo a povinnost vykonávat agresivní záchrannou funkci jako pojistný mechanismus konzervace digitálních informací v případě, že hrozí nebezpečí zničení nebo je daný materiál zanedbáván či opuštěn jeho současným opatrovníkem.*

(*Preserving digital information*, 1996)

Na základě těchto doporučení byly definovány další kroky na poli ochrany digitálních dokumentů. Ty budou aplikovány v rámci pilotních projektů, potřebných podpůrných struktur a tvorbou osvědčených postupů.

Tím vznikly základy pro spolupráci mezi informačními institucemi a výměnu know - how. Je nutné zajistit udržitelné financování a umožňovat sponzoring, konkurenční prostředí a volnou soutěž mezi vykonavateli zakázek. Digitální archivy by měly podporovat

praktické experimenty a demonstrace nových technologií (emulační software, migrační systémy, mechanismy autorizace) které rozvíjejí a usnadňují dlouhodobou ochranu digitální informace (Preserving digital information, 1996). V rámci podpůrných struktur je potřeba vytvořit informační infrastrukturu ve formě rozsáhlé sítě informačních profesionálů a propojených dat mezi institucemi. Tato síť je důležitá pro zajištění konzervace informace jako explicitního cíle. Pro další rozvoj je důležitý i dialog mezi vhodnými organizacemi ohledně standardizace a kritériích certifikace digitálních repositářů. Digitální repositáře by také měly mít schopnost agresivních zásahů v záchranné digitalizaci a stanovení administrativního bodu pro kontakt při koordinaci archivačních iniciativ.

Osvědčené postupy by měly podle Task Force vznikat v následujících oblastech: Design systémů umožňující dlouhodobou ochranu digitální informace, velkokapacitní úložiště kulturně cenných digitálních informací, požadavky a standardy popisu a správy digitálních informací a migrační postupy (Preserving digital information, 1996). Na základě dané analýzy je očekáváno od Komise skupiny vědeckých knihoven, že bude dané cíle následovat jak na národní, tak mezinárodní úrovni a bude podporovat vznik dialogu, interakce a produktů které pomůžou při vývoji důvěryhodných systémů a strategií (Preserving digital information, 1996).

1.1.2 Referenční rámec OAIS

OAIS (Open Archival Information System) je souhrn technických doporučení pro dlouhodobé, permanentní uchovávání digitálních dokumentů, který byl později vydán jako norma (ISO 14721:2003, 2003) a v roce 2012 byl rozšířen o připomínky z odborné komunity (ISO 14721:20012). Je dostupný i překlad normy v českém jazyce (HUTAŘ, 2015). Tento rámec nespecifikuje přesný design (Consultative committee, 2012) ani implementaci. Tento referenční model obsahuje:

- 1) *Rámec pro pochopení a zvýšení pozornosti o archivačních konceptech potřebných k dlouhodobé archivaci digitálních informací a přístupu k daným datům*
- 2) *Koncepty potřebné pro nearchivní organizace, aby se mohly plnohodnotně a efektivně zapojit do konzervačního procesu*
- 3) *Rámec, společně s terminologií a koncepty, potřebný pro popis a porovnávání architektury a operace současných a budoucích archivů.*
- 4) *Rámec popisující a porovnávající různé strategie a techniky pro dlouhodobou archivaci.*

- 5) *Základy pro porovnávání datových modelů digitální informace uchované archivy a diskutování o tom, jak se mohou datové modely dané informace v budoucnu změnit.*
- 6) *Rámce umožňující skrze další úsilí i zahrnutí ochrany informací, které nejsou digitální povahy (fyzická média a příklady)*
- 7) *Rozšíření konsenzu o částech a procesech dlouhodobé ochrany informací a přístupu k nim a podporování většího a konkurenčního trhu.*
- 8) *Příručky pro identifikaci a produkci OAIS příbuzným standardům*
(Consultative Committee, 2012)

1.1.2.1 Definice OAIS

V referenčním rámci je OAIS definován následovně:

„OAIS je archiv, skládající se z organizace (která může být součástí větší organizace), osob a systémů, která přijala odpovědnost za uchovávání informace a její zprostředkování určené komunitě. Splňuje soubor odpovědností definovaných v rámci sekce 4, umožňující OAIS Archivu odlišení od ostatních použití termínu „Archiv“. Výraz „otevřený“ v Otevřený archivační systém je použit pro implikaci, že tento Souhrn doporučení a jeho budoucí verze a standardy jsou vyvíjeny v otevřených fórech. Neznamena, že přístup do daného archivu je neomezený.“

(Consultative Committee, 2012)

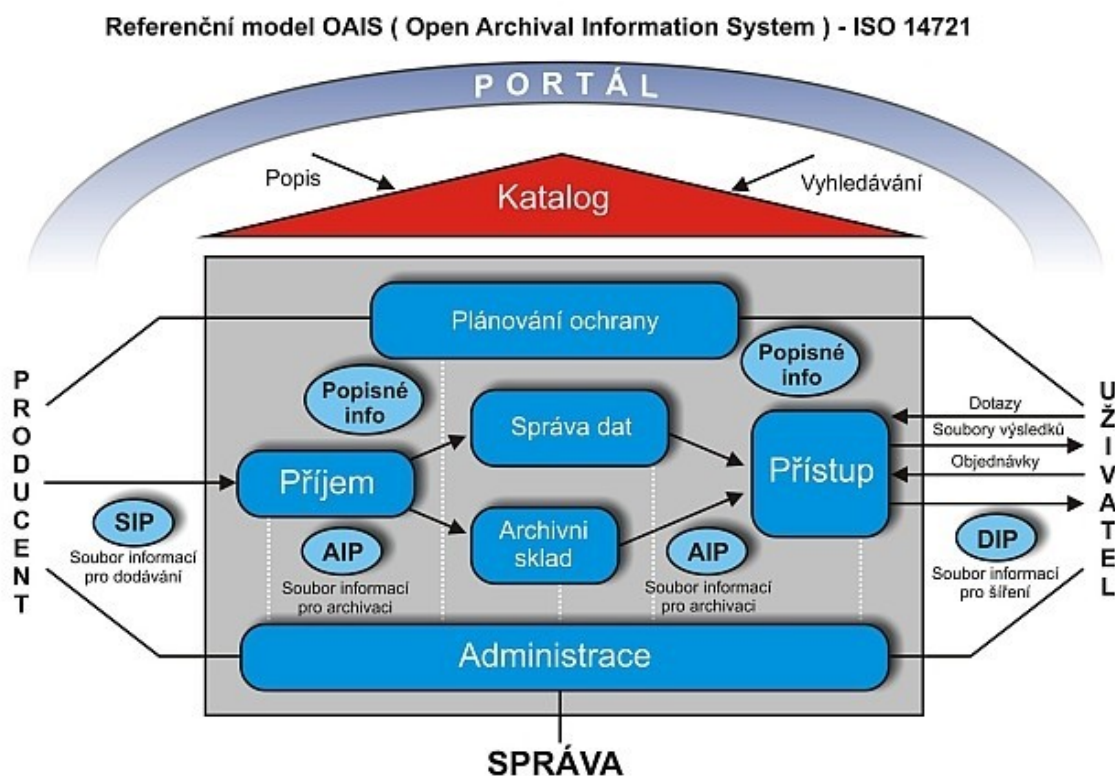
V rámci referenčního rámce je dále také definována široká škála termínů spojených s OAIS. Veskrze se jedná o snahu zamezit případným nezamýšleným významům daných pojmů.

1.1.2.2 Použitelnost a oblast působnosti OAIS

Referenční rámec lze aplikovat na prakticky jakýkoliv archiv, či instituce, které se zabývají dlouhodobou ochranou digitálních dokumentů. Dále je možné, že otevřený archivní informační systém nemůže nabízet služby přesahující jeho rámec. Předpokládá se, že implementátor OAIS nastaví služby tak, aby vyhovovaly instituci a jejím uživatelům. Lze tedy říci, že nevyžaduje specifické výpočetní platformy, systémovou strukturu nebo databázový design.

1.1.2.3 Koncepty OAIS a funkční entity

Na obrázku 1. je zobrazeno schéma OAIS (Rylich, 2010) a postup zpracování přijaté informace. Celkem je popsáno šest funkčních entit příbuzná rozhraní a vztahy mezi nimi. Pro zlepšení činnosti schématu budou v textu vyznačeny termíny ze schématu tučně a jejich překlady v závorkách. Svislé čáry pro administraci a plánování jsou přerušené pouze pro zlepšení orientace.



Obrázek 1 - Referenční model OAIS (Rylich, 2010)

Příjem poskytuje služby a funkce pro příjem dodavatelských balíčků (**SIP** – *submission information package*) od **Producenta** nebo z interní **administrace**. Entita pro příjem kromě samotného příjmu a kontroly kvality přetváří dodavatelské balíčky na archivní balíčky (**AIP** – *archival information package*), které připravuje na uložení v **Archivním skladu** a **Správu dat** uvnitř archivu podle jasně definované dokumentace a standardů.

Archivní sklad je služba pro ukládání, údržbu a vyhledávání archivních balíčků. Přijímá data z Příjmu a ukládá je do dlouhodobého úložiště. V rámci tohoto úložiště udržuje hierarchickou strukturu. Archivní sklad také zajišťuje udržování médií, na kterých jsou archivní balíčky uloženy a rutinní či namátkovou kontrolu chyb. Data archivace jsou následně zprostředkována pro Přístup. Archivace poskytuje návratový bod při nečekané události nebo katastrofě.

Správa dat zajišťuje popisné a administrativní informace objektů. Tyto metadata slouží k identifikaci dokumentů z Archivace a Administrace. Funkcemi datového managementu mohou být např. správa a aktualizace databází a jejich vnitřních definicí či referenční integrity (kontrola integrity vztahů mezi tabulkami nebo textem obsahujícími na sobě závislá data).

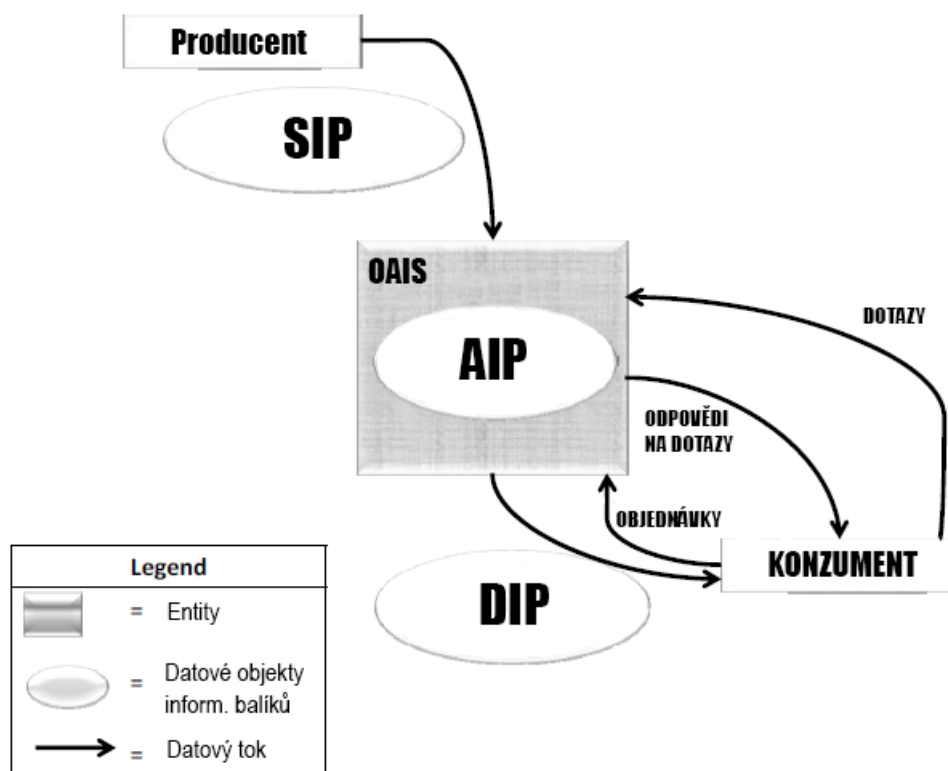
Administrace zajišťuje celkovou operaceschopnost OAIS. Tato entita zprostředkovává žádosti a vyjednávání dodavatelských smluv s Producenty. Provádí kontrolu přijímaných dokumentů ve formě dodavatelských balíčků (**SIP** – *submission information package*), aby bylo zajištěno, že splňují standardy archivu, který spravuje. Pod *administraci* také spadá správa, konstrukce a monitoring hardware a software systému, či podpora uživatelů a tvorba standardů.

Plánování dlouhodobé ochrany (není na obrázku, je součástí administrace a Archivního skladu) monitoruje prostředí OAIS, poskytuje doporučení a tvoří plány dlouhodobé ochrany, aby zajistila dostupnost a srozumitelnost informace pro cílovou skupinu. Konzervační plánování zahrnuje hodnocení obsahu archivu, navrhování aktualizací a migrací či migračních plánů a doporučení k úpravě dokumentace a standardům archivu. Dlouhodobé uchovávání by mělo být funkční i v případě, že se zprostředkovací technika stane zastaralou. Plánování dlouhodobé ochrany také vytváří šablony pro balíčky SIP, AIP a DIP a zajišťuje implementaci administračních cílů.

Přístup poskytuje služby pro zákazníky zjišťováním existence, popisu, lokace a dostupnosti informace uložené uvnitř OAIS. Zákazníci tedy mohou, pokud splňují podmínky přístupu žádat a dostávat informační produkty.

1.1.2.4 Prostředí OAIS

Prostředí otevřeného archivačního informačního systému je rozděleno na Producenty, Konzumenty a Management (Consultative Committee, 2012).



Obrázek 2 - Datový tok a prostředí OAIS

Producentem se rozumí fyzické nebo právnické osoby dodávající informace určené pro uchování. **Management** představuje instituce provozující strategii a vedení OAIS. Často je hlavní zdroj financování, který určuje využívání zdrojů jako je personál či vybavení a posuzuje risky spojené s administrací. Spolu s vedením OAIS lze očekávat ze strany Managementu k hodnocení jeho činnosti nebo plnění dlouhodobých plánů a interakci s ostatními entitami v prostředí OAIS. Rolí **Konzumenta** rozumíme osoby interagující s OAIS službami, popř. jevící zájem o uchovávané informace. Je také zmiňována cílová komunita, tedy okruh uživatelů schopných dané informaci porozumět. Individuální fyzická nebo právnická osoba může plnit roli jak Producenta, tak Konzumenta.

OAIS může v daném prostředí také například spolupracovat s jinými archivy na základě dohod mezi Managementy či přímo mezi OAIS na různých úrovních formality. Zároveň může docházet i k vytváření vztahů v rámci hierarchie. Například jeden OAIS může

fungovat jako Management nebo Konzument jiného otevřeného archivačního systému. To se může stát například v případech, kdy se Management rozhodne, že uchovávání méně využívané množiny dokumentů není třeba provádět lokálně.

1.1.2.5 Interpretace informace v OAIS

Hlavním atributem informace pro OAIS je její archivovatelnost. Datové objekty, interpretované zprostředkovanou informací tvoří tzv. informační objekt. Každý informační objekt by měl kromě obsahových informací obsahovat několik dalších typů informací.

Referenční informace identifikuje a v případě potřeby popisuje mechanismy, které byly použity při tvorbě obsahové informace objektu.

Kontextová informace vyjadřuje vztahy mezi obsahem objektů a jejich prostředím. Zároveň obsahuje informace o svém vzniku a jaké má objekt vztahy s ostatními informačními objekty v okolí.

Informace o původu je vlastně z části kontextová informace. Dokumentuje historii objekt a původ nebo zdroj obsahové informace včetně změn v obsahu a vlastnictví které objekt ovlivnily od doby jeho vzniku. To je zejména důležité pro ujištění uživatele o autenticitě dokumentu.

Informace o fixitě umožňuje kontrolu integrity dat pro zajištění, že objekt nebyl nijak změněn skrze nezdokumentované postupy. To zahrnuje kódování a detekční schémata jako prostředek pro zabránění takovýmto akcím.

Dalším typem je **informace o přístupu**, ta poskytuje údaje o restrikcích pro daný objekt, zahrnuje legální rámec, licence a kontrolu přístupu.

1.1.2.6 Zprostředkování informačních objektů

Informační objekt jako celek je třeba prezentovat takovým způsobem, aby ji mohl uživatel na základě jeho znalostní základny správně pochopit. Pro příklad by se daly například uvést cizí jazyky. V případě, že znalostní základna uživatelské skupiny např. neobsahuje znalost angličtiny a je třeba zprostředkovat kromě dané informace i slovník angličtiny. Zároveň je nutné počítat s tím, že se charakter uživatelské základny, tak její znalostní základny může kdykoliv změnit (Consultative Comittee, 2012).

1.1.2.7 Odpovědnost a povinnosti O AIS

Přímo v rámci standardu jsou definovány základní odpovědnosti O AIS:

O AIS bude:

1) Vyjednávat příjem a přijímat příslušné informace od informačních producentů

Pod tímto heslem si můžeme představit akvizici nových objektů (např. knihy, mapy, dokumenty šedé literatury, datové sety) pro O AIS.

Předpokládá se, že organizace O AIS bude provádět vyjednávání a příjem na základě interně stanovených podmínek. Ať už je to z její vlastní iniciativy, či je k tomu povinna. Příkladem daných kritérií může být unikátnost objektu, jeho zdroj, materiál či povaha technik použitých při jeho výrobě. Kromě povahy přijímaných objektů můžeme rozlišovat i způsoby předávání. Ty nejčastěji probíhají pomocí komunikačních cest jako jsou síť, email či specializované zásilky. O AIS i producent by měli být schopni zajistit dostatečné množství popisných informací (metadat) pro uživatelskou skupinu.

2) Získávat dostatečnou kontrolu poskytnuté informace na potřebné úrovni pro dlouhodobou archivaci.

U objektů je důležité vnímat možné rozdíly mezi fyzickým a duševním vlastnictvím či autorskými právy týkajícími se obsahové informace. V případě že je samotný O AIS je tvůrcem objektu a zároveň je vlastníkem autorských práv, má „volnou ruku“ pro potřebné procesy k dlouhodobé archivaci informace a jejímu zpřístupnění. V případě získávání obsahové informace od jakéhokoliv jiného producenta či entity je povinností O AIS zajistit právně platný přenos informace a jejího vlastnictví, nebo jasně vyjednat podmínky a omezení jejího užití. Ve většině případů je výhodné pro O AIS zajistit smluvní dohodu specifikující dané limity a umožňuje O AIS jednat bez nutnosti aktivního posuzování jednotlivých případů vlastníkem autorských práv.

To se může dít celkem ve třech kategoriích:

a) Smlouvy řešící autorská práva, duševní vlastnictví a legální omezení pro používání.

V případě autorských práv OAIIS musí znát a respektovat všechny legislativní aspekty dané problematiky. Je zde možnost i vytvoření speciálních pokynů pro příjem dokumentů a jejich šíření.

b) Oprávnění k modifikaci reprezentované informace

Přestože by se správně informace v případě dlouhodobé archivace modifikovat neměla. Může tato situace nastat při jejich zprostředkování, kdy je sice uchována ve své přesné formě, ale její prezentace není pohodlná pro uživatele. OAIIS tedy potřebuje legální i prostředkové možnosti pro migraci informací do nových prezentačních forem pro uživatelskou komunitu. To také může vyžadovat zapojení externích odborníků na dané téma, aby nebyla ohrožena ztráta informací.

c) Dohody s externími organizacemi

Dobrým příkladem takovýchto dohod může být spolupráce mezi OAIIS umožňující jedné instituci neuchovávat duplicitní základní informace a pouze na ně odkazovat do méně specializovaného OAIIS. Tyto úmluvy by měly být monitorovány pro kontrolu jejich užitečnosti a dodržování jejich plnění.

3) Stanovovat (ať už sám či ve spolupráci s ostatními organizacemi) vhodné cílové skupiny. Tím je také předpokládáno, že OAIIS bude schopen pochopit poskytnutou informaci a tím dané komunitě určit i znalostní základnu.

Pokud má informační systém zprostředkovávat obsahovou informaci, je nutné určit, komu. Tento postup hraje důležitou roli v OAIIS, protože pomáhá organizaci provozující systém určit, zda bude zprostředkovávaná informace uživateli či komunitě srozumitelná. Opět zde platí, že uživatelská komunita se může velice rychle změnit a možná bude potřeba dopředu předpokládat možné zúžení nebo rozšíření výběru zprostředkovaných dat.

4) *Zajišťovat, aby poskytnutá informace byla zachována „nezávisle pochopitelná“ pro cílovou skupinu. Dále také by cílová skupina měla mít možnost danou informaci pochopit bez nutnosti speciálních prostředků, jako například asistence expertů, který informaci vytvořili.*

Zprostředkovávaná informace by měla být dostatečně kompletní na to, aby ji mohla cílová komunita bez problémů interpretovat a využívat. Pro příklad můžeme uvést data z pozorování v průběhu vědeckého experimentu, který je potřeba přiblížit veřejnosti. Originální dokument v tomto případě může být srozumitelný pouze pro autory a specialisty oboru nebo rovnou čitelný pouze strojově. Pro danou cílovou skupinu pak nelze zprostředkovávat jenom čistá data z jednotlivých fází experimentu, ale je nutné přidat i vysvětlivky a doplnit informace o kontextu a významu daného dokumentu. Takto přidané informace je poté udržovat, doplňovat či upravovat v souladu s aktuální povahou neustále se měnící cílové skupiny.

5) *Dodržovat zdokumentované strategie a postupy, což zajistí že informace je uchována i v případě nepředvídané události včetně kolapsu archivu. Uchovaná informace by nikdy neměla být smazána bez toho, aby se tak jednalo v rámci schválené strategie. Nemělo by se mazat přes ad-hoc.*

Jednotný postup při tvorbě a údržbě archivních balíčků umožňuje zachování informací i nezávisle na jejich místě vzniku. Zároveň je možné data interpretovat různými způsoby skrze balíčky určené k šíření pro různé cílové skupiny, aniž by byla ohrožena integrita dat původního archivních balíčku. Zprostředkovávané balíčky by nicméně měly obsahovat informaci o tom, jakým způsobem byly vytvořeny z archivního balíčku.

6) *Umožňovat, aby byla uchovaná informace dostupná cílové skupině. OAIS by měl umožnit vytváření kopií jako obrazů původního objektu se zpětnou prokazatelností autenticity.*

Může existovat více možných způsobů, jak přistupovat k informacím poskytovaným OAIS. Jak již bylo krátce popsáno v odstavci výše, produktem OAIS pro dlouhodobou informaci jsou tzv. archivní balíčky (AIP – archival information package). Z nich jsou skrze zdokumentované postupy vytvářeny uživatelské balíčky (DIP – dissemination information package).

1.1.2.8 Dlouhodobá ochrana a její perspektivy v rámci OAIS

Referenční rámec také popisuje různé postupy, jaké byly nebo mohou být použity při dlouhodobé ochraně. Bez ohledu na to, jak dobře OAIS spravuje svoje aktuální sbírky, časem bude třeba migrovat média na nové nosiče. To může zahrnovat kopírování, upravování nebo přidávání obsahu nebo archivních informací na nové médium, úpravu nebo přidání software a hardware důležitého pro obsahovou informaci a její uchování či úpravu struktury archivního balíčku. K tomu slouží proces **digitální migrace**. Ta je definována jako: *“Přenos digitální informace za účelem jejího uchování v rámci OAIS.”* (Consultative Committee, 2012). Oproti normálním transferům se liší důrazem na archivaci obsahové informace jako celku a vnímáním nové archivní implementace jako rovnocennou náhražkou té staré. Celá kontrola a odpovědnost všech aspektů je kompletně v rukou OAIS. Takovýmito transferům bychom měli v tomto poli rozumět na širší úrovni, tzn. jakýmkoliv změnám v bitech nebo obsahové informaci apod. Zároveň je také digitální informace jeden z procesů, při kterých je větší riziko ztráty dat a je proto k ní přistupovat s velkou mírou pečlivosti.

Digitální migrace jako taková může mít různé motivátory. Patří zde například zlepšení efektivity a snížení nákladů, nebo nové požadavky ze strany služeb uživatelům a zastarávání nosičů. Díky rychlému vývoji nových technologií a softwaru může OAIS získat větší úložiště za stejnou nebo nižší cenu, nebo zrychlit fungování svých interních procesů. Zlepšování designu balíčků může zajistit jejich větší nezávislost na nosiči a tím dopomoci k pohodlnější a méně náročné migraci. Uživatelská komunita v rámci rozvoje technologií také zvyšuje své nároky a očekávání na služby OAIS. To vyžaduje úpravu uživatelských balíčků nebo přizpůsobování obsahu archivních balíčků pro cílovou uživatelskou skupinu. Dále je také potřeba vnímat skutečnost že se digitální média v poslední době stávají méně spolehlivými pro bezpečné uchovávání informací. I přestože jsou nosiče používány s opatrností, je nutné je po čase měnit.

OAIS rozlišuje 4 typy migrace digitální informace:

(seřazeno podle vzrůstajícího riziku ztráty dat)

1. *Operace nevyžadující změny bitových sekvencí*

a. *Renovační migrace*

Typ digitální migrace, kde instance nosiče, obsahující jeden nebo více archivních balíčků je nahrazena instancí jiného nosiče stejného typu pomocí kopie média použitého pro původní AIP a jeho správu či přístup. Výsledkem tohoto procesu je archivní balíček, který je stále možné bez jakékoliv změny v archivní infrastruktuře dále identifikovat nebo k němu přistupovat. Tento typ je nejméně riskantní, protože se žádné bity obsažené v původním balíčku, které prezentují jeho obsah, nebo umožňují jeho vyhledávání a přístup nemění. Příklad renovace může být náhrada poškozeného CD-ROMu za nový, obsahující naprosto identické informace.

b. *Duplikační migrace*

Digitální migrace, při které se nezasahuje do obsahové, strukturní či popisné informace. Bity jsou v nezměněné podobě přesunuty na stejný nebo nový typ nosiče. Jedná se svým způsobem o typ obnovy, nicméně místo jedné jednotky vzniknou dvě stejné jednotky na stejném či různých nosičích. Duplikace může vyžadovat změny do mapovací struktury archivních jednotek. Při replikaci dochází pouze k minimální úpravě na bitové úrovni, a to pouze v případě že se replikuje na jiný typ nosiče.

2. *Operace vyžadující změny bitových frekvencí*

a. *Balíčkovací migrace*

Digitální migrace se změnou v informacích o balíčku. Při tomto procesu se až na výjimky nezasahuje do obsahové informace, ale je už podstatnější riziko nějaké chyby v procesu migrace než u předchozích dvou. Pro představu můžeme mít CD-ROM obsahující tři soubory, jeden obsahující metadata, druhý obsahovou informaci a třetí informace o struktuře balíčku. Obsah celého CD je přesunut na nový nosič a změní se implementace adresářů a souborů. Tento typ migrace je například potřeba při změně způsobu správy dat.

b. Transformace

Typ digitální migrace se změnou v obsahové nebo popisné informaci se snahou zachovat plnohodnotnou obsahovou informaci, přestože se změní prezentace balíčku. Jedná se o nejriskantnější způsob migrace. Přetransformovaný archivní balíček je brán jako **nová verze** původního balíčku a originál může být po nějakou dobu uchován pro ověření kvality zachované obsahové informaci. Transformaci také můžeme rozlišovat na **reverzibilní** a **ireverzibilní**.

1.2 *Standardy v České republice*

Dne 5.12. 2016 byla schválena Národní koncepce dlouhodobé ochrany digitálních dat v knihovnách (Ústřední knihovnická rada ČR, 2016), která vychází z Koncepce rozvoje knihoven ČR na léta 2011–2015 (Ministerstvo Kultury, 2012).

Jejím účelem je připravit v národním měřítku podmínky pro uchovávání a dlouhodobou ochranu digitálních dat na základě standardu NDK a systému pro trvalou identifikaci Uniform Resource Name: National Bibliography Number (URN:NBN) (Kvašová, 2017).

V rámci schválené koncepce byly stanoveny následující cíle a principy:

- *Rozvoj a plánování dlouhodobé ochrany*
- *Zlepšení organizačních podmínek*
- *Podpora spolupráce a sdílení*
- *Snížení bariér*
- *Rozvoj mechanismů financování dlouhodobé ochrany*
- *Používání (zahraničních modelů pro stanovení nákladů na dlouhodobou ochranu*
- *Podpora vzdělávání, kvalifikace a výzkumu*
- *Metodické centrum pro dlouhodobou ochranu digitálních dat*
- *Důraz na prosazování dlouhodobé ochrany*

(Ústřední knihovnická rada ČR, 2016)

Do roku 2020 se předpokládá že vznikne garantované datové centrum pro zajištění bitové ochrany digitálních dat knihoven a LTP systém pro zajištění logické ochrany, popř. lokální řešení jednotlivých institucí různé úrovně a kvality. Procesu dlouhodobé ochrany se budou účastnit metodické centrum, centrální datové centrum a jednotlivé knihovny s využitím stávajících (Visk7 apod.) i nových finančních mechanismů. Budou schváleny standardy pro data a metadata, doporučení a metodiky pro dlouhodobou ochranu na základě mezinárodních norem (Ústřední knihovnická rada ČR, 2016).

1.2.1 Dlouhodobá ochrana digitálních dokumentů

Plně vyvinutou standardizaci v ČR budeme schopni hodnotit až okolo roku 2020 po splnění cílů z Národní koncepce dlouhodobé ochrany digitálních dokumentů. Pro potřeby této práce jsou místo toho prezentovány myšlenky z knihy Ladislava Cubra – Dlouhodobá archivace digitálních dat.

Text se snaží o konceptualizaci a hlubší porozumění problematiky v oblasti dlouhodobé archivace vzhledem k tomu že české prostředí v těchto oblastech poněkud zaostává (Cubr, 2010). Z textu byly vybrány 2 kapitoly, které jsou důležité pro případovou studii v praktické části práce. Jedná se o Překážky dlouhodobé archivace a Ochranná opatření digitální archivace.

1.2.1.1 Překážky dlouhodobé ochrany

Text v této kapitole popisuje několik rovin, které mohou působit problémy v dlouhodobé ochraně digitálních informací. Jedná se o technologickou, informační, systémovou a institucionální rovinu. Každá z těchto rovin popisuje různá ohrožení a překážky pro dlouhodobou ochranu digitálních dokumentů (Cubr, 2010).

1.2.1.1.1 Technická rovina

Nosičům digitální informace hrozí kromě fyzické a chemické degradace i jejich zastarávání (Hedstrom, 1997). Degradace nemusí nutně probíhat jenom v dlouhodobém horizontu, ale může jít i o jednorázové selhání jako např. výpadek proudu, selhání zálohování, hackerské útoky a viry, nebo lidský faktor. Je tedy nutné se snažit podobné události předvídat a snažit se těmto rizikům zabránit předem (vícenásobné zálohy, adekvátní zabezpečení atp.) tím že budeme vybírat kvalitní, více stabilní prostředky a materiály. Vhodné je také podrobně sledovat technologický vývoj a dostatečně aktualizovat software a hardware digitálního repositáře.

1.2.1.1.2 Informační rovina

Pokud jsme schopni zachovat informace v dostatečném technickém stavu, je důležité abychom byli schopni chránit i jejich čitelnost a srozumitelnost. Na této rovině dlouhodobou archivaci ohrožují zejména formátová rizika. Prvním z dvou největších rizik je zastarávání formátu.

Tyto změny je podobně jako trh s novými technologiemi sledovat a adekvátně na ně reagovat. Ukončení podpory formátu může být podle Cubra (2010) sice někdy výhodou, protože to umožňuje více programům data interpretovat, nicméně časem povede k jeho kompletnímu zániku, protože se formát nebude přizpůsobovat uživatelským a technologickým nárokům, které budoucnost přinese. Kromě aktualizace používaných formátů je také třeba vybírat více robustní řešení, která jsou více odolná proti degradaci dat a poškození bitových sekvencí.

Dalším problémem v této rovině je, že digitální objekty nejsou samy o sobě schopny auto reference a samy toho o sobě moc nevypráví. Je zapotřebí dokumentace, která daný formát a jeho fungování popisuje a vysvětluje. Specifickou formou dokumentace jsou i metadata.

1.2.1.1.3 Systémová rovina

Nestačí chránit samotné digitální objekty ale i sbírky, ve kterých jsou organizovány. Počet digitálních dokumentů neustále roste. To zvyšuje i nároky na jejich správu v rámci repositáře. Velké množství souborů snižuje přehlednost a prodlužuje vyhledávání, digitální objekty je potřeba jednoznačně identifikovat pomocí systému trvalých identifikátorů jako např. URN: NBN (jmenný prostor National Bibliographic Name pro Uniform Resource Name) nebo DOI (Digital Object Identifier). Digitální repositář musí také zajistit ochranu a udržování práv digitálního objektu. Kromě práv jako závazků pro repositář je dobré vědět, že legislativa digitálních práv může dlouhodobou archivaci komplikovat (De Beer, 2016).

Digitální dokumenty je mnohem snazší změnit než jejich fyzické protějšky, kontrola a prevence porušení integrity souborů nebo celých balíčků je tedy prioritou pro uchování plnohodnotné obsahové informace. S integritou také souvisí snaha zachovat autenticitu souborů jednak jako potvrzení jejich důvěryhodnosti, tak i umožnění původního uživatelského zážitku. Autentické dokumenty lze považovat za nejvyšší cíl pro dlouhodobou archivaci (Caplan, 2008).

1.2.1.1.4 Institucionální rovina

Tato rovina řeší digitální repositáře v rámci organizačního uspořádání. Ve většině případů nejsou samostatné, ale spíše součástí větší organizace ve formě poskytované služby. To ovlivňuje celou řadu aspektů od ekonomických záležitostí až po nabízené služby či obsah fondu. Digitální repositář jako takový by se měl snažit o co největší míru důvěryhodnosti (Gladney, 2007). Dále má administrativní odpovědnost, organizační životaschopnost,

je finančně udržitelný, systematicky bezpečný a technologicky a procesně odpovědný (RLG, OCLC, 2002).

Podobně jako většina paměťových institucí musí digitální repositáře brát v potaz ekonomická rizika a opatrně nakládat s finančními prostředky. Je nutné také počítat s organizačními a procesními překážkami. Oproti klasickým předmětům archivace mají digitální dokumenty mnohem kratší ochranný horizont a liší se i v akvizici. Zatímco dříve bylo možné archivovat vše, dnes musí, vzhledem k objemu dat, instituce vybírat co je pro dlouhodobou ochranu vhodné a co ne. Digitální dokumenty také prochází změnou v oblasti jejich vlastnictví, v dnešní době paměťové instituce často nemají přímo kopii, ale pouze přístup, což ztěžuje archivaci.

1.2.1.2 Ochranná opatření dlouhodobé archivace

Zajištění ochrany dlouhodobé archivace může probíhat několika způsoby, od strategií a technik či metod po doporučení nejvhodnějších postupů. Text zde popisuje nejdůležitější metody, jak řešit překážky z přechodí kapitoly v rámci osvědčených postupů.

1.2.1.2.1 Ochranná opatření technologické roviny

Na technologické rovině je hlavním cílem zabránit ztrátě dat z důsledku selhání hardware, popř. jejich nečitelnosti z důvodu zastarání technologií. Prvním často aplikovaným opatřením je výběr vhodných nosičů. Vzhledem k předchozí praxi se už dále nedoporučuje používat optické nosiče, respektive používání takovýchto nosičů je vzhledem k jejich povaze už spíše riziko (Bradley, 2006). Jako lepší alternativa se jeví zejména magnetické nosiče, protože v současné době exponenciálně roste jejich kapacita a klesá cena (Gladney, 2007). Dalšími výhodami pevných disků je snadná automatizace, testování chyb a odhalování blížící se degradace nosiče.

V rámci digitální migrace se repositář snaží zamezit nebo neprovádí změnu bitových sekvencí čímž není kompromitována integrita a autenticita digitálních dat. Nicméně tato metoda zaručuje pouze přežití fyzické formy digitálního objektu. V delším období je často aplikovaným postupem technologická migrace podle konceptu OAIS. Ta vyžaduje režim pravidelných přesunů digitálních objektů na novější nosiče a technologická prostředí podle režimu stanoveného repositářem. Větší digitální repositáře často nemají data uložena na pouze jedné lokaci, ale vytvářejí se kopie na různých místech např. v jiné části světa či státu. Tím se předchází ztrátě dat v případě postižení lokace přírodní katastrofou. Dále se dříve objevovaly kromě analogového zálohování i snahy o konzervaci technologie.

Vzhledem k tempu, jakým je ukončována podpora zastaralého hardware, a i samotná životnost je relativně velmi krátká je tento postup označován jako neproveditelný (Cubr, 2007). Pro záchranu digitálních objektů je možné použít takzvanou „digitální archeologii“. Problémem tohoto opatření je jeho poměrně vysoká nákladovost a je tedy vnímáno spíše jako krizové řešení než opatření pro dlouhodobou archivaci.

1.2.1.2.2 Ochranná opatření informační roviny

Ochrana na této úrovni bojuje zejména vůči zastarávání formátů a problematice jejich specifikace. První praktikou bývá tzv. zapouzdření digitálních objektů. To v praxi znamená že kromě samotného média a popisné informace je v pouzdru obsažen i např. software umožňující trvalé zpřístupnění digitálního objektu. Je tím ale ztížena možnost formátové migrace v případě zastarávání formátu nebo přibaleného software a také hrozí že v budoucnosti nebude existovat hardware schopný software spustit. Tomu by měla napomoci strategie Digitální Rosettské desky (Robertson, 1996). V rámci této strategie není archivována znalost přirozeného jazyka ale fungování technologického prostředí. Toho je dosaženo přesnou a kompletní dokumentací původního prostředí na jejíž základě by jej bylo možné v budoucnosti rekonstruovat. Názory na toto opatření, jeho efektivitu a proveditelnost se ale liší. K otázkám rekonstrukce a zprostředkování původního prostředí se dnes nejčastěji přistupuje pomocí emulace, což je ve své podstatě znovuvyvoření původního software a hardware starší platformy na jiné, modernější platformě. Speciálním typem je poté emulace modulární, která používá pro každou komponentu vlastní emulátor. který s ostatními komunikuje a spolupracuje.

V případě formátů tak se uplatňují různé strategie. Vždy je důležité posoudit vlastnosti formátu podle jeho vlastností pro dlouhodobou archivaci z hlediska jeho dostupnosti a formě dokumentace. Pro účely dlouhodobé archivace rozlišujeme formáty proprietární a neproprietární, které mohou mít otevřenou nebo uzavřenou dokumentaci. Pro účely dlouhodobé archivace se nejvíce hodí neproprietární formáty s otevřenou dokumentací, přičemž naprosto nepoužitelné jsou formáty s uzavřenou dokumentací (ať už proprietární nebo neproprietární). Dalšími kritérii jsou poté míra standardizace formátu, exportovatelnost, závislost na platformách a komplexita.

V případě zastarávání formátu se přistupuje se k formátové migraci, při které jsou bitové sekvence uspořádány podle nové struktury, přičemž nedojde ke změně obsahu. Tato změna bývá často ireversibilní, je tedy nutné ji v případě digitálních repositářů pečlivě naplánovat a bezchybně provést. Častým případem, kdy dochází k formátové migraci mimo

zastarávání formátu je formátová normalizace probíhající například při vytváření dodavatelských balíků na vstupu archivačního procesu. Tento přístup ale nelze udržovat věčně. Přistupuje se tedy k pragmatičtějšímu řešení, kdy jsou nové formáty, vyvíjeny organizacemi jako W3C nebo OASIS přizpůsobovány tak, aby splňovaly požadavky na jejich interoperabilitu a nebyly omezovány právy duševního vlastnictví.

Dále vznikají formátové registry – služby poskytující technické údaje o formátech, jejich identifikaci, validaci a charakterizaci.

Kromě samotných objektů prochází modifikacemi a standardizací i jejich metadata. Existuje několik různých typologií, nicméně pro účel této práce bude použita typologie a popis funkcí citované přímo v knize:

- Administrativní metadata sloužící ke správě a administraci dat
- Archivační metadata pro řízení ochrany dokumentu
- Technická metadata popisující funkce systémů a chování metadat
- Uživatelská metadata používaná pro monitoring úrovně a formy užívání dat

(Baca, 2008)

Metadata obecně plní zejména tyto funkce:

- Popis jednotek sbírky
- Podpora vyhledávání a identifikace
- Seskupování jednotek v repozitáři
- Záznam důkazu autenticity
- Podpora integrity jednotek
- Záznam přístupových omezení
- Podpora výměny informací mezi autonomními repozitáři
- Záznam technických informací

(Gladney, 2007)

Deskriptivní, administrativní a strukturální metadata jsou nejčastěji kódována podle schématu METS. Výstupem tohoto standardu jsou poté METS (*Metadata Encoding and Transmission Standard*) xml dokumenty (kontejnery) obsahující informace pohromadě. Tímto způsobem mohou sloužit jako součásti dodavatelských, archivačních a uživatelských balíků. Dalším významným standardem je PREMIS (*PREservation Metadata: Implementation Strategies*), který se soustředí přímo na archivační metadata pro knihovny.

Do nejčastěji využívaných ochranných opatření v této skupině patří metadatová migrace. Ta veskrze probíhá stejně jako formátová migrace a slouží k úpravě, obohacování metadat a dokumentaci nových ochranných opatření.

1.2.1.2.3 Ochranná opatření systémové roviny

Na systémové rovině jsou nejčastějšími tématy identifikace dokumentů, digitální práva, integrita a autenticita při velkém počtu dokumentů. Každý repositář vyžaduje kromě hardware a obsahu také příslušný software disponující nástroji pro interpretaci či úpravu uložených dat. Jejich příkladem by mohly být opensource projekty jako DSpace a Fedora Commons, či komerční systémy jako Rosetta, Tesella SDB (Safety Deposit Box) nebo DIAS (Digital information archiving solution) od IBM. Jak komerční, tak nekomerční systémy splňují některé požadavky normy OAIS, nicméně na kompletní řešení se stále čeká.

V současnosti se z řešení nejvíce osvědčila tzv. infrastruktura trvalé identifikace (Cubr, 2007). Tímto ochranným opatřením rozumíme přidělování trvalých, jedinečných (ve vztahu identifikátor – resolver) a spolehlivých identifikátorů. Tím je lze umožnit dlouhodobé zpřístupňování dokumentu, konzistentní výsledky vyhledávání i v případech, kdy je dokument přesunut z jednoho místa na druhé a snazší management jak samotných objektů, tak i digitálních práv. K tomu nejlépe slouží globální identifikační systémy. V současné době existuje hned několik vhodných identifikátorů jako třeba Handle, DOI, a URN: NBN (používaný i NTK). Identifikátory jsou přidělovány digitálním objektům skrze registry identifikátorů. V daném registru jsou pak obsaženy ještě další metadata a informace o aktuálním umístění těchto objektů. Společně s registry jsou většinou také provozovány resolvers, které tyto informace zprostředkovávají z přístupných digitálních repositářů.

Spolehlivé ochrany digitálních dokumentů lze na této úrovni docílit skrze sledování některých principů. Prvním je správa digitálních práv. Jedná o kontrolu přístupu uživatelů na základě různých pravidel stanovených legislativou a repositářem. To také souvisí s kontrolou autenticity dokumentů, protože je-li narušena identita nebo integrita dokumentů, tak není jenom ohrožena archivace dokumentu, ale může zároveň dojít i k narušení práv. Přístup nelze zprostředkovávat centrálně, ale spíše skrze síť na sobě nezávislých, ale propojených repositářů (tzn. distribuovaný přístup) s použitím gridových technologií. To vše probíhá za soustavné autentifikace kontrolující integritu a autenticitu dokumentu, prokazující provenienci a spravující verze.

1.2.1.2.4 Ochranná opatření institucionální roviny

Na institucionální rovině jsou zajišťovány otázky důvěryhodnosti repositáře, ekonomická rizika a překážky v rámci instituce. Důvěryhodnost zajišťuje certifikace digitálního repositáře (Cubr, 2007). V současné době existuje hned několik certifikačních nástrojů, které zjišťují okolnosti jeho fungování na základě připravených otázek a příkladů. V současnosti jsou nejznámější nástroje TRAC (*Trustworthy Repositories Audit & Certification*) Nestor (*Network of Expertise in Long-Term Storage of Digital Resources*) a Drambora (*Digital Repository Audit Method Based on Risk Assessment*). Udržitelnost repositáře je zajišťována pomocí ekonomických modelů. Ty umožňují monitoring a správu instituce a její efektivní financování. Důležitou roli při financování hrají motivační mechanismy, které přitahují potenciální stakeholdery (vydavatele, producenty a zástupce státního sektoru rozhodující o financování). Nicméně vzhledem k tomu, že každý stakeholder má různé zájmy, je nutné pro ně vytvářet i specifické motivační mechanismy. Pro správné fungování repositáře je důležitá také existence obchodních modelů, podle kterých probíhá, pokud možno, co nejužší spolupráce s dodavateli.

Obchodní modely a motivační strategie se ukazují také jako výhodné v rámci komplexního plánování, při kterých je vybíráno z jednotlivých opatření a je vytvářena dlouhodobá strategie. Plánovat lze jak pro již stávající repositáře, za pomoci auditů a certifikace, tak pro nově vznikající. Základní rámec pro nově vznikající repositáře lze převzít z referenčního rámce OAIS, nicméně další postup je lepší plánovat pomocí konkrétnějších nástrojů. V textu je jako příklad uveden PLATTER (*PLAnning Tool for Trusted Electronic Repositories*). S jeho pomocí jsou stanovovány funkce a účel repositáře, objem ukládaných dat, počty zaměstnanců, definice procedur a technické implementace. Na základě těchto rozhodnutí je pak formována dlouhodobá strategie a dílčí plány pro konkrétní oblasti.

2 Audit digitální knihovny Národní technické knihovny

Tato kapitola popisuje samotný průběh auditu digitální knihovny a jeho výstupy.

2.1 Metodika auditu

Zvýšením důvěryhodnosti digitálního repositáře umožňujeme lepší poskytování služeb uživatelské komunitě. V rámci auditu digitální knihovny bylo vybíráno z několika nástrojů na základě poznatků získaných studiem dokumentů z teoretické části této práce.

Norma OAIS zmiňuje další doporučené standardy, které vycházejí z referenčního rámce. Jsou jimi ISO 9000 a příbuzné standardy pro kontrolu kvality, ISO 17799:2005 pro bezpečnost dat a management informací, ISO 15489:2001 pro systematický přístup k dokumentům a norma ISO 14721:2003 pro Otevřený archivační informační systém (Consultative Committee, 2012).

Výběr byl na konec zúžen na 2 hlavní nástroje, které jsou pro digitální knihovny používány: ISO 16363 (TRAC) a Drambora. Oba zmíněné umožňují tzv. self-audit a jsou mezinárodně uznávané. Oba nástroje byly v minulosti opakovaně použity i v České Republice např. v Národní knihovně a ÚVT Univerzity Karlovy v Praze (Hutař, 2008) nebo přímo v rámci NTK pro Úložiště šedé literatury (NUŠL, 2011).

Jako vhodnější byla nakonec zvolena norma ISO 16363, jako jediná z výběru má svůj český normativní překlad: ČSN ISO 16363 (Úřad pro technickou normalizaci, metrologii a státní zkušebnictví, 2014)

2.2 Cíle auditu

Hlavním cílem auditu je popsat aktuální stav digitální knihovny NTK z hlediska dlouhodobé ochrany digitálních dokumentů. V současné době probíhá přípravná fáze projektu implementujícího úložiště, na které budou digitální objekty ukládány pro dlouhodobou ochranu a audit by měl dopomoci při vyhledávání nedostatků, které je potřeba vyřešit před uvedením úložiště do provozu. Zároveň by měl tento audit poskytnout personálu spravujícímu digitální knihovnu poznatky pro její budoucí rozvoj.

Audit digitální knihovny NTK popisovaný v rámci práce se skládá z následujících fází:

1. Testování důvěryhodnosti digitálního repositáře vnitřním auditem podle normy ISO 16363.
2. Posouzení silných a slabých stránek digitální knihovny v oblastech organizace, správy digitálních objektů a technické infrastruktury. Identifikace největších rizik
3. Formulace krátkodobých a dlouhodobých cílů pro zlepšení důvěryhodnosti a funkčnosti digitálního repositáře.

2.3 *Aktuální stav digitální knihovny Kramerius NTK*

Digitální knihovna NTK má několik specifíků, kterými se liší od digitálních repositářů určených přímo k archivaci digitálních dat.

Digitalizáty se v první řadě vytváří pro usnadnění přístupu k historickému fondu a preventivní ochraně dokumentů. Může nastat i situace, kdy bude zapotřebí záchranná digitalizace. Dlouhodobou archivaci digitálních dokumentů lze tedy považovat jako zálohu zálohy, vzhledem k tomu že existují dokumenty historického fondu uložené ve skladech, ty jsou zpřístupňovány kromě výpůjček ve studovně starých tisků i skrze digitální knihovnu a tyto digitální objekty jsou následně archivovány v nově vznikajícím dlouhodobém úložišti. Národní Technická knihovna nepatří mezi knihovny s právem povinného výtisku a nemá pro své digitální objekty legislativně danou archivační povinnost. V rámci koncepce uchovávání digitálních dokumentů je na tuto situaci obecně nahlíženo takto:

Knihovní zákon výslovně nehovoří o knihovních fondech a dokumentech v digitální podobě. Z definice pojmů knihovní fond a knihovní dokument v § 2 knihovního zákona lze ale usuzovat, že digitální podoba je v nich zahrnuta. Podobná situace panuje i v dalších paměťových institucích. V oblasti archivů např. archivní zákon (Zákon o archivnictví a spisové službě a o změně některých zákonů č. 499/2004 ve znění novely 167/2012) vymezuje v § 2 bodě b) pojem péče o archiválie, který vedle výběru, evidence, zpracování zahrnuje také ochranu archiválií.

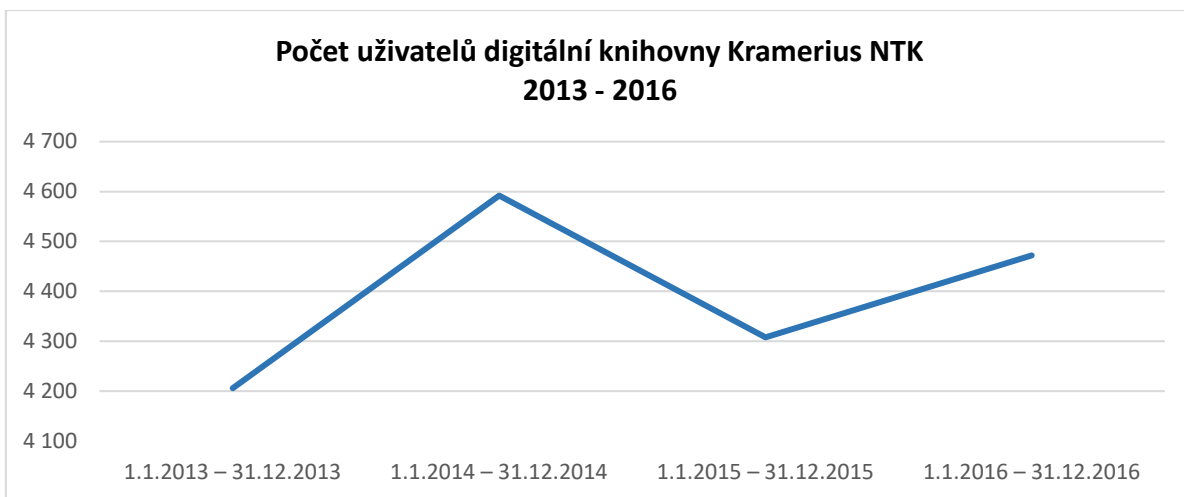
Z obou zákonů tedy vyplývá, že se knihovny a archivy musejí zabývat i ochranou dokumentů, které existují a do institucí přicházejí pouze v digitální podobě (bez fyzického protějšku).

(Ústřední knihovnická rada ČR, 2016)

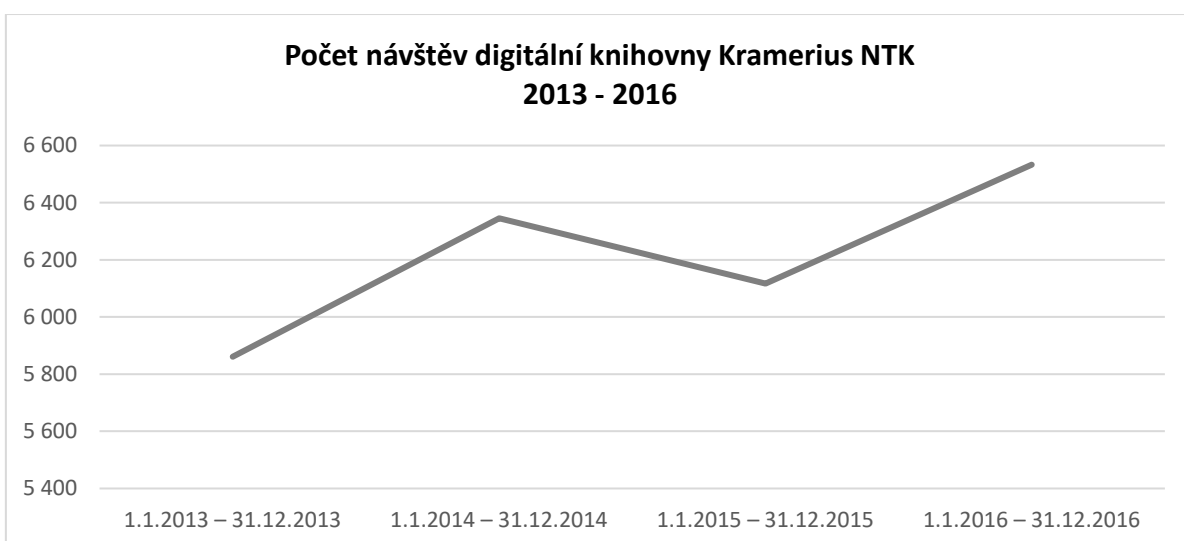
Až do poměrně blízké doby byly u digitálních objektů uchovávány na úložišti pouze skeny. V případě odstranění objektu z digitální knihovny bylo potřeba znovu vyhledat skeny na úložišti, celý objekt opětovně popsat a nahrát zpět na Kramerius. Nikdy naštěstí nedošlo k nějakému většímu selhání, kdy by se ztratilo velké množství dokumentů ale v případě,

že by k takovému selhání někdy došlo by obsah začínal opět od nuly (Jan Kolátor, 2017). Ten samý scénář by se dal popsat i pro úložiště skenů, tzn. v případě, že by došlo nedopatřením k smazání záložních skenů, bylo by potřeba dokumenty opětovně zdigitalizovat, což by jednak bylo náročné pro samotný historický fond, tak pro personál. Kramerius NTK již skoro půl roku úspěšně testuje a používá digitalizační nástroj ProArc (Production and Archival system), který poskytuje širokou škálu nástrojů pro jednak zpřístupňování, tak dlouhodobou archivaci digitálních dat. Skeny jsou vytvářeny z části od externího dodavatele a také personálem digitalizačního oddělení Národní technické knihovny. Lze tedy říct že je knihovna sama sobě z části dodavatelem, což usnadňuje komunikaci a dokumentaci zpracovacího postupu. Oddělení digitalizace také provádí u některých dokumentů kontrolu výstupu rozpoznávání textu, vzhledem k tomu že tato technologie stále není dokonalá. Administrátor digitální knihovny skeny kontroluje a případně opravuje a poté z nich pomocí ProArcu vytváří NDK-PSP balíky obsahující uživatelské a archivní kopie, technická a obsahová metadata a údaje o provenienci a jednoznačný identifikátor URN: NBN. V případě externí digitalizace jsou dodávány už hotové balíky, které prochází jenom kontrolou. Vytvořené NDK-PSP balíky jsou poté zpřístupňovány skrze digitální knihovnu Kramerius. Ten je schopen z nich interpretovat skeny a metadata potřebná pro uživatele a není tedy potřeba vytvářet speciální zprostředkovatelské balíky. Nezměněné balíky jsou také odeslány pro dlouhodobou archivaci na nově vyvíjené úložiště LTP. Druhou částí povinností administrátora je poté vývoj a správa digitální knihovny po technické stránce komunikace s uživateli.

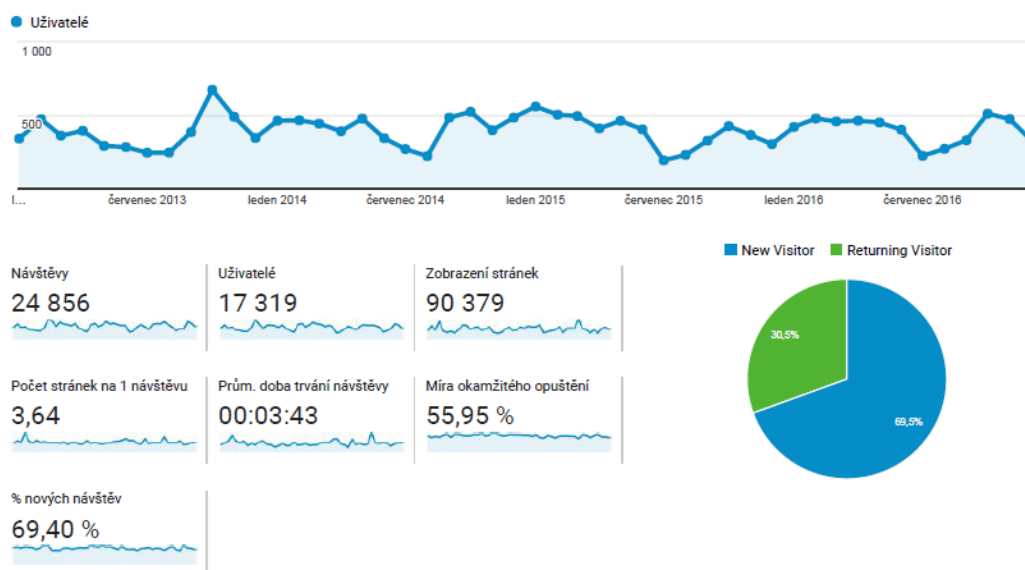
Samotná digitální knihovna jako taková má podle statistik sledovaných přes Google Analytics stabilní uživatelskou základnu (viz. Graf 1, Graf 2) a nezaznamenává žádné razantní výkyvy počtu návštěvníků mimo letní období, kdy digitální knihovnu navštěvuje méně studentů. (viz. Graf 3, podrobněji Příloha 2 – Výstup Google Analytics digitální knihovny NTK 2013-2016)



Graf 1 - Počet uživatelů digitální knihovny NTK 2013–2016



Graf 2 - Návštěvnost digitální knihovny NTK 2013–2016



Graf 3 - Souhrn návštěvnosti digitální knihovny NTK 2013–2016

2.4 *Audit ISO podle ISO 16363*

V rámci zpřehlednění bylo upraveno číslování kapitol v práci tak, aby odpovídalo použitému číslování v rámci normy ISO 16363¹. Jednotlivé splněné či nesplněné požadavky jsou vyznačeny kurzivou. V některých případech může být více požadavků odpovězeno v rámci jednoho výroku.

Po prostudování normy probíhal audit získáním informací z patřičné dokumentace a komunikací s Mgr. Janem Kolátorem za oddělení ICT Služeb a Ing. Ondřejem Kochem z oddělení ICT infrastruktury NTK. Na základě těchto poznatků poté bylo posuzováno, zda digitální knihovna NTK splňuje či nesplňuje jednotlivé požadavky dané normou.

3. Organizační infrastruktura

3.1. Vedení a životaschopnost organizace

3.1.1. Úložiště musí mít poslání, které vyjadřuje závazek uchovávat, dlouhodobě udržovat, spravovat a zpřístupňovat digitální informace.

Existují texty na webu knihovny popisující účel, fungování knihovny a její obsah. Nicméně faktor dlouhodobé archivace není zmíněn (Digitální knihovna, 2017). *Cílová skupina z textu vyplývá*, nicméně by bylo vhodné ji přímo definovat a doplnit. Text na webu je také zobrazován ve zkrácené podobě přímo v Krameriovi, nicméně není dostatečně obsáhlý a má veskrze pouze stručně informativní charakter.

Digitální knihovna NTK požadavek **ISO 16363 v bodě 3.1.1 splňuje pouze minimálně** a je nutné prohlášení zpracovat.

3.1.2. Úložiště musí mít strategicky plán uchování, který stanoví, jak úložiště zajistí, že své poslání bude vykonávat dlouhodobě.

V současné době takovýto *plán neexistuje*, dlouhodobá ochrana dokumentů je prozatím přítomna pouze ve formě praktik zálohování skenů a serverových záloh. Pracuje se na spuštění a provozu úložiště pro dlouhodobou archivaci.

Pro případy ukončení provozování digitální knihovny ze strany NTK nebo jejího zřizovatele, Ministerstva školství, mládeže a tělovýchovy ČR, *není zpracován žádný plán nástupnictví*. V současné době by se taková situace řešila ad hoc, zřejmě zároveň s celkovými opatřeními při rušení nebo významné změně instituce vlastníci rozsáhlé knihovní, papírové i digitální, fondy.

¹ Kapitola 3 odpovídá sekci 3 v normě. (pozn. autora)

V rámci organizačního uspořádání NTK je správa digitální knihovny NTK zařazena v Odboru pro informační technologie, ICT služby a infrastrukturu. *Vedoucí tohoto odboru je přímo podřízen řediteli instituce, účastní se porad vedení a o případné kritické situaci související např. se zrušením instituce by měl být bezprostředně informován.* Kritická situace by nastala pouze tehdy, pokud by o něčem takovém nebyl informován ani ředitel instituce.

Digitální knihovna NTK v rámci **ISO 16363 v bodě 3.1.2 splňuje 1 ze 3 požadavků.**

3.1.3. Úložiště musí mít pravidla pro získávání obsahu nebo jiný dokument upřesňující typ informací, které bude úložiště uchovávat, udržovat, spravovat a zpřístupňovat.

Fond Digitální knihovny NTK je vytvářen na základě jednotlivých digitalizačních projektů řešených NTK od roku 2008 (Digitální knihovna, 2017). První verze Strategie digitalizace byla zpracována v roce 2015 a předložena vedení NTK. Porada vedení vrátila dokument k dopracování, k čemuž dodnes nedošlo (Národní technická knihovna, 2016).

Digitální knihovna NTK požadavek **ISO 16363 v bodě 3.1.3 nesplňuje** a je nutné strategii dopracovat a schválit.

3.2. Organizační struktura a personální zajištění

3.2.1. Úložiště musí mít stanoveny, které úkoly má plnit, a musí zaměstnávat pracovníky, kteří pro jejich plnění mají potřebné dovednosti a zkušenosti.

Pracovní povinnosti jsou jasně definovány v rámci *pracovní smlouvy zaměstnance*. V současné době je přímo pro digitální knihovnu zaměstnán pouze administrátor, který provádí správu, včetně vytváření a oprav metadat pro vstupní balíky a údržbu. Zaměstnanec se také *podílí na vývoji* Digitální knihovny v rámci účasti na schůzkách pracovního týmu pro vývoj SW Krameria. Administrátor se také *pravidelně účastní na školeních v rámci i mimo instituci.* (Jan Kolátor, 2017)

Digitální knihovna NTK v rámci **ISO 16363 v bodě 3.2.1 splňuje 3 z 3 požadavků**

3.3. Procesní zodpovědnost a rámec pravidel pro uchovávání

3.3.1. *Úložiště musí mít vymezenou svou určenou skupinu a související znalostní základnu (základny) a její vymezení musí být vhodným způsobem zpřístupněno.*

Definice cílové skupiny a příslušné znalostní základny jsou vymezeny, vycházejí z knihovního řádu (Národní technická knihovna, 2017). Další možné omezení je třeba vnímat v limitech copyrightu, kdy jsou některé dokumenty přístupné pouze z budovy knihovny (viz. Řízení přístupu).

Digitální knihovna NTK v bodě **3.3.1 splňuje** požadavek normy ISO 16363.

3.3.2. *Za účelem plnění strategického plánu uchovávání musí mít úložiště zavedena pravidla pro uchovávání.*

Daná pravidla v digitální knihovně NTK v dostatečné míře neexistují. V současné době zatím neexistuje v České Republice softwarové open-source řešení, které by umožňovalo tyto pravidla aplikovat. V letech 2016-2020 probíhá vývoj nástroje ArcLib jako komplexního řešení pro dlouhodobou archivaci digitálních sbírek (Melichar, 2016). V cílech tohoto projektu je i vytvoření metodiky a řešení pro fyzické ukládání dat a zajištění bit-level ochrany. NTK zvažuje využití této metodiky. Prozatím jsou tyto zásady v digitální knihovně zatím zdokumentovány pouze v rámci Redmine, systému sloužícímu k evidenci úkolů a vedení provozní dokumentace. Tato dokumentace je přístupná pouze zaměstnancům a má charakter pracovních poznámek. Do roku 2020 se jí jako finančně neefektivní vyvíjet nebo zakoupit systém, který by tyto požadavky splňoval.

V předchozích letech pravidelný audit Digitální knihovny NTK neprobíhal. Doporučená periodicita je roční (Hruška, 2013). Zavádění auditů digitální knihovny NTK je pouze v rovině úvah. Konkrétní úkol nebyl zadán žádnému pracovníkovi. Nicméně ochota podstoupit audit je signálem, že o tuto aktivitu NTK zájem má.

Digitální knihovna NTK požadavek **ISO 16363 v bodě 3.3.2 nesplňuje**.

3.3.3. *Úložiště musí mít zdokumentované provedené změny týkající se provozu, postupů, softwaru a hardwaru.*

V současné době je vývoj a popis digitální knihovny dokumentován v rámci systému pro flexibilní management projektů Redmine (<http://www.redmine.org/>). Dále jsou veškeré změny v konfiguracích Krameria a editoru metadat ProArc na serveru přímo v souborech *okomentovány datem, jménem pracovníka, co změnu provedl a krátkým popisem proč a co udělal*. Tyto informace jsou poté dostupné pracovníkům ICT s dostatečnými přístupovými

právy. V současné době také probíhá migrace vývoje modifikací pro Digitální knihovnu NTK na veřejně přístupný repositář na Github.com, vytvořený speciálně pro Digitální knihovnu NTK. Aktualizace je součástí pracovní náplně administrátora. *Změny hardware jsou dokumentovány oddělením ICT infrastruktury* v rámci na základě kupních smluv a interní dokumentace v systému Redmine.

Digitální knihovna NTK požadavek **ISO 16363 v bodě 3.3.3 splňuje.**

3.3.4. *Úložiště se musí zavázat k otevřenosti a zodpovědnosti za veškerá opatření, jež podporují provoz a správu úložiště a mají vliv na uchovávání digitálního obsahu v průběhu času.*

Většina procesů a postupů je v rámci NTK *dostupná na webu knihovny* (Digitalizační workflow, 2017), popřípadě detailněji *na požádání*. Výjimkou jsou licencovaná softwarová řešení, nebo o informace které by mohly poškodit bezpečnost repositáře.

Digitální knihovna NTK požadavek normy **ISO 16363 v bodě 3.3.4 splňuje.**

3.3.5. *Úložiště musí stanovit způsob měření integrity informací, získávat výsledky z měření, sledovat je a vhodně je vyjadřovat.*

NTK zatím nemá k dispozici program pro kontrolu integrity digitální knihovny a jejího obsahu prostřednictvím pravidelných auditů obsahu. V současné době se přešlo na tvorbu NDK-PSP balíků pro usnadnění archivace. V časovém horizontu přibližně tří měsíců bude implementována kontrola integrity balíků pomocí validátoru vyvíjeným Národní knihovnou, který umožňuje kontrolu struktury a integrity, validace metadat pomocí šablon a validací obrazových dat a poskytne textový výstup o zpracování a chybách (Komplexní validátor NDK, 2017). Tyto informace poté lze vybraným způsobem zprostředkovávat.

NTK **bude** požadavek **ISO 16363 v bodě 3.3.5 splňovat** po implementaci výše zmíněného validátoru.

3.3.6. *Úložiště se musí zavázat k plánu pravidelného sebehodnocení a externí certifikace.*

V NTK je tento *self-audit prvním pokusem*, o významu auditu je NTK informována a záleží na rozhodnutí vedení, zda se bude v těchto auditech pokračovat. (Jan Kolátor, 2017)

Digitální knihovna NTK v požadavek **ISO 16363 v bodě 3.3.6 nesplňuje.**

3.4. Ekonomická udržitelnost

3.4.1. *Úložiště musí mít zavedeny krátkodobé i dlouhodobé procesy podnikového plánování pro udržení provozu úložiště v průběhu času.*

Financování a plánování zajišťuje instituce NTK. Digitální knihovna NTK jako taková nemá vlastní rozpočet a je financována jako součást oddělení ICT služeb.

Digitální knihovna NTK v rámci bodu 3.4 ISO 16363 splňuje 3 ze 3 požadavků.

3.5. Smlouvy, licence a právní odpovědnost

3.5.1. *Úložiště musí mít a udržovat patřičné smlouvy nebo dohody o vkládání dat vztahující se k digitálním dokumentům, které spravuje, uchovává a/nebo zpřístupňuje.*

Hardware využívaný pro digitální knihovnu NTK je majetkem instituce, software (Kramerius, Proarc, perspektivně ArcLib) jsou českými open-source systémy vyvíjenými v rámci projektů podporovanými státní správou. Veškerým digitálním objektům ukládaným do digitální knihovny NTK má knihovna práva na základě knihovní licence autorského zákona (Zákon o právu autorském, 2000, § 37).

Národní technická knihovna si nechala vypracovat právní posudek a nyní posuzuje akvizici, zpřístupňování a uchovávání práv na základě tohoto dokumentu (Polčák, 2009).

NTK po celou dobu existence Digitální knihovny neřešila případ reakce na závazek. Na webových stránkách Krameria a NTK jsou dostupné kontakty na odpovědné osoby. Případná oprava zpřístupnění na základě stížnosti by byla okamžitě opravena administrátorem.

Digitální knihovna NTK v rámci ISO 16363 v bodě 3.5.1 splňuje 4 z 4 požadavků.

3.5.2. *Úložiště musí sledovat a spravovat práva duševního vlastnictví a omezení užití obsahu úložiště v souladu s požadavky dohody o vkládání dat, smlouvy nebo licence.*

Knihovna opět vychází ze znaleckého posudku pro digitalizaci (Polčák, 2009). Dále jsou aktivně sledovány proměny s proměnami autorského zákona. Dokumenty jsou rozdělovány na „veřejné“ a „privátní“.

Veřejné dokumenty jsou většinou dokumenty, na které se autorská práva buď nevztahují, popřípadě už zanikly (uběhla doba 70 let od smrti autora). V rámci Krameria je i aplikována pohyblivá zeď, která automaticky do této skupiny zařazuje dokumenty, u kterých je už jisté, že už mohou být veřejně přístupné (uběhlo 100 let od vydání).

Dokumenty stále spadající pod ochranu jsou přístupné pouze z budovy NTK, nicméně je knihovna může dlouhodobě archivovat, vzhledem k tomu že dlouhodobé úložiště dat není přístupné veřejnosti.

Digitální knihovna NTK požadavek normy **ISO 1636 v bodě 3.52 splňuje**.

4. Správa digitálních objektů

4.1. Příjem: získávání obsahu

4.1.1. Úložiště musí stanovit informační obsah a vlastnosti informací, které bude uchovávat.

Strategický plán dlouhodobého uchovávání *zpracován není*. Existuje podrobný popis digitalizačního workflow (Řihák, 2014), *jehož verze prozatím nebyla aktualizována tak, aby přesně reflektovala aktuální stav*. Tento dokument a provozní dokumentace by měly posloužit jako základ pro aktualizovanou verzi práce se systémem ProArc a následně i pro definici informačního obsahu a vlastností informací v rámci strategie digitalizace a dlouhodobého uchovávání.

Pro typologii objektů ukládaných v NTK je používána typologie MARC21 (Marc standards, 2017; Národní technická knihovna, 2016) ukládaná v katalogizačních záznamech Národní technické knihovny. Tyto parametry jsou použity pro definici obsahových informací. Tento požadavek není splněn u části dokumentů uložených v NTK před implementací postupů přebírání metadat z katalogu instituce. To souvisí s vývojem katalogizace v ČR.

Digitální knihovna NTK v rámci **ISO 16363 v bodě 4.1.1 nesplňuje žádný ze tří požadavků**. Nicméně disponuje prostředky na jejich splnění.

4.1.2. Úložiště musí jasně určit, které informace je s konkrétním informačním obsahem třeba propojit v okamžiku jeho příjmu.

Digitální knihovna NTK využívá pro příjem informací strukturu NDK-PSP balíků, které mají jasně definovanou strukturu a obsahují formáty definované v rámci mezinárodních standardů, jako jsou METS, MODS, Dublin Core, ALTO apod.

Digitální knihovna NTK požadavek normy **ISO 16363 v bodě 4.1.2 splňuje**.

4.1.3. Úložiště musí mít odpovídající popisy, které mu umožní rozpoznat a zpracovat balíčky SIP.

Přijímané externě digitalizované balíky jsou prozatím kontrolovány pouze administrátorem digitální knihovny na základě popisu struktury balíků popsané na základě specifikace NDK (Národní digitální knihovna, 2017), v budoucnosti se bude v rámci standardizace a automatizace používat aktuálně zaváděný validátor NDK-PSP.

Digitální knihovna NTK požadavek normy **ISO 16363 v bodě 4.1.3 nesplňuje** do doby zprovoznění validátoru NDK.

4.1.4. Úložiště musí umět patřičně ověřit totožnost tvůrců všech dokumentů.

V rámci NDK-PSP balíků obsahuje každý SIP informace o provenienci a tvůrci balíku (Standardy digitalizace NDK, 2017). V případě externí digitalizace je také dodáván seznam s URN: NBN identifikátory.

Digitální knihovna NTK požadavek normy **ISO 16363 v bodě 4.1.4 splňuje**.

4.1.5. Úložiště musí mít proces příjmu, který ověří úplnost a bezchybnost každého balíčku SIP.

Balíky vytvářené přímo v rámci NTK jsou ověřovány na výstupu z metadatového editoru. Bez správné struktury a obsahu balíku nelze přiřadit URN: NBN identifikátor a tím dokončit proces vytváření (Standardy digitalizace NDK, 2017). U externí digitalizace budou balíky ještě procházet skrze validátor. Všechny balíky budou také validovány před vložením do dlouhodobého úložiště.

Digitální knihovna NTK požadavek normy **ISO 16363 v bodě 4.1.5 nesplňuje** do doby zprovoznění validátoru NDK.

4.1.6. Aby úložiště mohlo digitální objekty uchovávat, musí získat možnost s nimi dostatečně nakládat.

Každý spuštěný proces avizuje svůj stav zpracování přímo v rozhraní systému Kramerius. Z hlediska administrace je toto informování dostatečné.

Digitální knihovna NTK požadavek normy **ISO 16363 v bodě 4.1.6 splňuje**.

4.1.7. Úložiště musí v dohodnutých bodech procesu příjmu poskytovat tvůrci/vkladateli přiměřenou odezvu.

Oddělení digitalizace zapojuje administrátora digitální knihovny do výběru dokumentů k digitalizaci, dále je poskytována zpětná vazba v průběhu digitalizace a kontrola během nahrávání do digitální knihovny NTK. Zprostředkovatel externí digitalizace komunikuje s administrátorem digitální knihovny podobným způsobem. Pracovní postup a pracovní postupy jsou dokumentovány v rámci smlouvy (viz. Příloha 1 Smlouva o digitalizaci dokumentů NTK)

Digitální knihovna NTK požadavek normy **ISO 16363 v bodě 4.1.7 splňuje.**

4.1.8. Úložiště musí mít aktuální záznamy o opatřeních a administrativních procesech, které jsou potřebné pro získávání obsahu.

V administrátorském rozhraní systému Krameria a v popisné informaci balíčků jsou dostupné údaje o uložení. NTK dále také externím poskytovatelům odesílá potvrzení o přijetí (Jan Kolátor, 2017). Digitální knihovna NTK požadavek normy **ISO 16363 v bodě 4.1.8 splňuje.**

4.2. Příjem: vytváření balíčku AIP

4.2.1. Úložiště musí ke každému balíčku AIP nebo ke každé třídě balíčků AIP, které uchovává, mít přiřazené jeho vymezení, které je vhodné pro zpracování balíčku AIP a která splňuje potřeby dlouhodobého uchovávání.

Všechny balíčky jsou vymežovány v rámci zjednodušení archivačního procesu jako archivační, nerozčleňují se a jsou zcela identické s balíčky zpřístupňovanými uživatelům. Digitální knihovna NTK ale nemůže vzhledem k nedostatečné dokumentaci prokázat použití svého vymezení při získávání informačního obsahu a PDI.

Digitální knihovna NTK v rámci **ISO 16363 v bodu 4.2.1 splňuje 2 ze 3 požadavků.**

4.2.2. Úložiště musí mít popsán způsob tvorby balíčků AIP z balíčků SIP.

Tento proces je popsán v rámci dokumentace NDK (Standardy digitalizace NDK, 2017).

Digitální knihovna NTK požadavek normy **ISO 16363 v bodě 4.2.2 splňuje.**

4.2.3. *Úložiště musí dokumentovat, jakým způsobem ve výsledku zpracovalo balíčky SIP.*

Tato dokumentace v současné době neexistuje. V budoucnu bude v NTK součástí strategie dlouhodobé ochrany digitálních informací.

Digitální knihovna NTK požadavek normy **ISO 16363 v bodě 4.2.3 nesplňuje.**

4.2.4. *Úložiště musí mít a užívat pravidla, na jejichž základě jsou vytvářeny trvalé a jedinečné identifikátory balíčků AIP.*

Digitální knihovna NTK společně se systémem ProArc používá pro identifikaci dokumentů mezinárodní *identifikátor URN: NBN*. Ten je *jedinečný* jak v rámci úložiště, tak globálně a je přidělován všem vznikajícím balíčkům. Pro všechny dokumenty vytvořené před implementací ProArcu *existuje identifikátor UUID (universally unique identifier)*, který je v rámci úložiště *jedinečný*. Nicméně jako ideální stav se předpokládá, že všechny dokumenty budou přítomny v NDK-PSP balíčku a budou vybaveny URN:NBN (Jan Kolátor, 2017).

Digitální knihovna NTK *může poskytnout kompletní seznam* všech URN:NBN a UUID identifikátorů, ale v případě UUID *neprovádí namátkové kontroly jedinečnosti*. V případě nejedinečnosti identifikátoru by byl ale administrátor upozorněn už při procesu vytváření digitálního dokumentu. Systém URN:NBN je používán jak v ČR i ve světě knihovnami uchovávajícími mnohem větší množství dokumentů. *Nelze tedy předpokládat, že by NTK mohla v blízké době dosáhnout jeho kapacity* (Jan Kolátor, 2017).

Digitální knihovna NTK je také propojena s katalogem NTK, kde jsou uvedeny jednoznačné odkazy na digitální knihovnu.

Digitální knihovna NTK v rámci **ISO 16363 v bodě 4.2.4 splňuje 5 z 6 požadavků.**

4.2.5. *Úložiště musí mít přístup k nástrojům a zdrojům, které jsou třeba k tomu, aby pro všechny digitální objekty mohly být poskytovány směrodatné vysvětlující informace.*

Digitální knihovna NTK disponuje nástroji a metodami pro zjištění formátu souborů u všech dodaných datových objektů. Administrátor digitální knihovny podle dohody s oddělením digitalizace určuje, zda je třeba dodat další vysvětlující informace pro daný objekt v rámci příloh, které jsou s dokumentem pevně spojeny v rámci struktury NDK-PSP balíku (Jan Kolátor, 2017).

Digitální knihovna NTK v rámci **ISO 16363 v bodě 4.2.5 splňuje 4 ze 4 požadavků.**

4.2.6. *Úložiště musí mít zdokumentované procesy pro získávání informací o uchovávání (PDI) vztahující se k informačnímu obsahu a získávat PDI v souladu se zdokumentovanými procesy.*

Běžné provozní postupy a digitalizační workflow jsou obsaženy v provozní dokumentaci systému Redmine NTK a na webových stránkách (Digitalizační workflow NTK, 2017). Nicméně neexistuje zobrazitelná dokumentace způsobu, jakým úložiště získává a spravuje informace o uchovávání (Jan Kolátor, 2017).

Digitální knihovna NTK v **ISO 16363 nesplňuje žádný z 3 požadavků v bodě 4.2.6.**

4.2.7. *Úložiště musí zajistit, že v okamžiku vytvoření balíčků AIP je určené skupině jejich informační obsah srozumitelný.*

Digitální knihovna NTK spravuje pouze dokumenty, které se nacházejí přímo v jejím fondu, lze tedy předpokládat jejich srozumitelnost pro definovanou cílovou skupinu (Jan Kolátor, 2017). Nicméně *neexistuje dokumentace* pro daný proces. *Nelze tedy otestovat srozumitelnost pro informační skupinu, nebo reagovat v případě že je příliš nízká* a tím pádem Digitální knihovna NTK v rámci **ISO 16363 v bodě 4.2.7 nesplňuje žádný ze 3 požadavků.**

4.2.8. *Úložiště musí úplnost a bezchybnost každého balíčku AIP ověřit v okamžiku vytvoření.*

Ze struktury NDK-PSP balíčku *nelze vypustit jeho označení identifikátorem*. V případě, že není digitální objekt kompletní (Standardy digitalizace NDK, 2017). Tento mechanismus je přímo implementován i v editoru metadat ProArc.

Digitální knihovna NTK požadavek **ISO 16363 v bodě 4.2.8 splňuje.**

4.2.9. *Úložiště musí stanovit nezávislou metodu ověření integrity sbírky / obsahu úložiště.*

Digitální knihovna NTK splňuje 7 z 10 podmínek dokumentace z bodů 4.2.1–4.2.4, které jsou pro tento bod podstatné. Metoda pro nezávislou kontrolu je stanovena, nicméně není zatím implementována (viz. bod 3.3.5 – Validátor NKP).

Digitální knihovna NTK požadavek **ISO 16363 v bodě 4.2.9. nesplňuje** do doby implementace validátoru.

4.2.10. Úložiště musí mít aktuální záznamy o opatřeních a administrativních procesech, které se vztahují k vytvoření balíčku AIP.

Tato dokumentace pro digitální knihovnu NTK neexistuje. Digitální knihovna NTK požadavek **ISO 16363 v bodě 4.2.10 nesplňuje.**

4.3. Plánování uchovávání

Oddíly 4.3. a 4.4. byly v rámci auditu záměrně vynechány, vzhledem k tomu že úložiště pro dlouhodobé uchovávání digitálních dokumentů teprve vzniká a audit má také účel získání patřičných dat a poznatků pro přípravu digitální knihovny na dlouhodobou archivaci.

Pro úplnost lze tedy předpokládat, že Digitální knihovna NTK v rámci **ISO 16363 v bodě 4.3 nesplňuje žádný ze 4. požadavků.**

4.4. Uchovávání balíčku AIP

Viz. bod 4.4.

Digitální knihovna NTK v rámci **ISO 16363 v bodě 4.4 nesplňuje žádný ze 4. požadavků.**

4.5. Správa informací

4.5.1. Úložiště musí stanovit základní požadavky na informace a tím určené skupině umožnit, aby dokumenty našla a zjistila, zda o ně má zájem.

Pro využívání digitální knihovny stačí internetový prohlížeč bez jakýchkoliv doplňků. Všechny objekty v digitální knihovně jsou alespoň minimálně popsány v rámci metadat MODS a DC. Tato informace bude doplněna do údajů na webu knihovny administrátorem digitální knihovny ve spolupráci s webmasterem a PR oddělením. Další informační požadavky budou specifikovány v rámci doplňování definice cílové skupiny pro digitální knihovnu.

Digitální knihovna NTK požadavek normy **ISO 16363 v bodě 4.5.1 bude splňovat** po doplnění informací na web knihovny.

4.5.2. *Úložiště musí získat nebo vytvořit základní popisné informace a musí zajistit jejich propojení s balíčkem AIP.*

Každý vytvořený archivační balík obsahuje převzatá metadata z knihovního katalogu a jednoznačný identifikátor umožňující jej identifikovat jak uvnitř, tak mimo instituci.

Digitální knihovna NTK požadavek normy **ISO 16363 v bodě 4.5.2 splňuje.**

4.5.3. *Úložiště musí udržovat obousměrnou vazbu mezi každým balíčkem AIP a k němu náležejícími popisnými informacemi.*

Viz. bod 4.5.4

4.5.4. *Úložiště musí v průběhu času udržovat propojení mezi balíčky AIP a popisnými informacemi.*

Tento bod je vyřešen přímo umístěním popisné informace přímo uvnitř archivního balíku. Mimo jiné lze balík také dohledat skrze resolver NKP a zobrazit metadata přímo na webu, popřípadě skrze systémové číslo dohledat informace v katalogu. Dále jsou také všechny digitální dokumenty označeny jednoznačným identifikátorem, popř. existuje dokument se seznamem zdrojů pro ověření existence digitální kopie, skrze který lze patřičná data dohledat, který je umístěn na webu knihovny (Digitální knihovna, 2017)

Digitální knihovna NTK normu **ISO 16363 v bodech 4.5.2 a 4.5.3 splňuje.**

4.6. Řízení přístupu

4.6.1. *Úložiště se musí řídit pravidly pro zpřístupňování.*

Tento bod je popsán přímo na webu (Digitální knihovna, 2017) tak přímo uvnitř digitální knihovny (Kramerius NTK, 2017). Popis obsahuje, co a za jakých podmínek je komu dostupné a jak jsou uživatelé autorizováni.

Do logu na serveru se pak zapisuje, odkud bylo k datům přistupováno a co přesně bylo požadováno v případě chyby zpřístupnění. Tyto logy jsou v pravidelných intervalech kontrolovány a administrátor digitální knihovny na ně reaguje.

Digitální knihovna NTK normu **ISO 16363 v bodě 4.6.1 splňuje v 2 ze 2 požadavků.**

4.6.2. *Úložiště se musí řídit pravidly a postupy, které umožňují poskytování digitálních objektů, u nichž lze zpětně dohledat původní dokumenty, s dokladem o jejich hodnověrnosti*

V současné době je jediná dostupná dokumentace pro systémový design na stránkách projektu Kramerius (Kramerius Wiki, 2017). Metadata digitálního objektu lze zobrazit přímo v rozhraní Krameria. Exportované PDF dokumenty sice neobsahují žádnou evidenci autenticity, nicméně je každý dokument opatřen persistentním URL, přes které lze originál zobrazit a porovnat jej s kopií.

Digitální knihovna NTK umožňuje dva druhy zpětné vazby. Jeden pro uživatele určený přímo knihově provozující digitální knihovnu, nebo pro knihovníky skrze Github, kde lze nahlašovat případné bugy nebo návrhy na vylepšení.

Digitální knihovna NTK normu **ISO 16363 v bodě 4.6.2 splňuje 2 z 2 požadavků.**

5. Infrastruktura a řízení bezpečnostních rizik

5.1. Řízení rizik technické infrastruktury

5.1.1. *Úložiště musí zjišťovat a řídit rizika infrastruktury systému vztahující se k provozu a cílům v oblasti uchovávání.*

Technická kontrola probíhá za pomoci oddělení ICT Infrastruktury NTK, *používá se důvěryhodný software* s velkou komunitou, konkrétně software jmenovaný i v rámci normy např. Apache Tomcat, SOLR, Fedora repository (ČSN ISO 16363, 2014).

Odhady životnosti hardware se posuzují na základě záruky od dodavatele (Ondřej Koch, 2017). Digitální knihovna NTK vzhledem k svojí velikosti a obsahu v současné době *neprovozuje systém pro sledování technologických novinek*, provádění postupů pro sledování a upozornění kdy hardwarové technologie zastarají nebo je třeba je změnit. *To samé se dá říci i o zaznamenávání a reagování na bezpečnostních rizika.* Veškeré tyto činnosti jsou manuálně prováděny administrátorem digitální knihovny za pomoci oddělení ICT infrastruktury (Ondřej Koch, 2017).

Nedochází k výpadkům kvůli přetížení, procesy vyžadující odstavení serveru jsou minimalizovány a prováděny mimo špičku (Jan Kolátor, 2017). Lze tedy říci, že Digitální knihovna NTK *je dobře vybavená pro svoji cílovou skupinu i nápor, který servery podstupují za běžného provozu.*

Vzhledem k rozsahu digitální knihovny *nedochází k záloze* obsahu digitální knihovny samostatně. V rámci infrastruktury pobíhá pouze záloha na magnetické pásky a personál je schopný obnovit hardware pouze v případě selhání dílčích částí. Nicméně v případě že selže diskové pole je obsah nenávratně ztracen.

Digitální knihovna NTK *má adekvátní monitoring* pomocí software, co zasílá upozornění administrátorům v případě abnormálního chování (docházející místo na disku, výkon CPU, výpadky).

Personál prochází školeními bezpečnosti práce vyplývajícími ze zákoníku práce. Pouze s rozlišením na vedoucí a zaměstnance.

Digitální knihovna NTK v rámci normy **ISO 16363 v bodě 5.1.1 splňuje 8 z 16 požadavků**. Tato situace je především daná škálou digitální knihovny, kvůli které nelze zajistit dostatečné financování infrastruktury a personálu, aby dosáhla úrovně certifikovaného úložiště, či certifikace v rámci infrastruktury a bezpečnosti.

5.1.2. Úložiště musí spravovat počet a umístění kopií všech digitálních dokumentů.

Digitální knihovna NTK na základě zpětné vazby uživatelů a činnosti administrátora *náhodně kontroluje dostupnost digitálních objektů* a jejich výskyt jak v úložišti, tak v rámci systému Krameria (Jan Kolátor, 2017). Synchronizace vícenásobných kopií se neprovádí, do příslušných balíčků se po jejich vytvoření nezasahuje. Digitální knihovna NTK proto tedy *nemá způsob ověření, zda jsou digitální kopie identické*. V budoucnosti by se tato činnost měla provádět na základě skriptu využívajícího validátor NDK-PSP balíčků (Jan Kolátor, 2017).

Digitální knihovna NTK v rámci normy **ISO 16363 v bodě 5.1.2 splňuje 1 z 2 požadavků**.

5.2. Řízení bezpečnostních rizik

5.2.1. Úložiště musí zajišťovat systematickou analýzu bezpečnostních rizik, která souvisí s daty, systémy, zaměstnanci a fyzickým zařízením.

Viz. bod 5.2.2.

5.2.2. Úložiště musí mít zaveden způsob řízení, který patřičně řeší každé z vymezených bezpečnostních rizik.

Digitální knihovna NTK *nesleduje postupy stanovené normami řady ISO 27000*, vzhledem k vybavení a škále *nelze předpokládat budoucí certifikaci* (Ondřej Koch, 2017). Neexistuje seznam pro řízení rizik. Vzhledem k charakteru sbírky je velmi nízká

pravděpodobnost, že se stanou předmětem útoku a jednotlivým dokumentům není přiřazován stupeň ochrany. Případné útoky lze zaznamenat a podniknout kroky proti jejich opakování, nicméně normou popsaná efektivní ochrana je v rámci digitální knihovny NTK *neproveditelná*.

Digitální knihovna NTK normu **ISO 16363 v bodech 5.2.1 a 5.2.2 nesplňuje**.

5.2.3. Zaměstnanci úložiště musí mít vymezené role, odpovědnosti a oprávnění vztahující se k provádění změn v systému.

V rámci digitální knihovny je zaměstnán pouze administrátor (Jan Kolátor, 2017). Ten disponuje superuživatelskými privilegii ke všem systémům souvisejícím s digitální knihovnou. V případě práce stážistů na odborné praxi je používán dočasný účet s omezenými privilegii. Digitální knihovna NTK nepoužívá postupy stanovené v normách řady ISO 2700 a nemá speciální dokumentaci oprávnění k systému, vzhledem ke škále a financování je to neproveditelné (Ondřej Koch, 2017).

Digitální knihovna NTK normu **ISO 16363 v bodě 5.2.3 nesplňuje**.

5.2.4. Úložiště musí mít vhodné písemně zaznamenané plány připravenosti na živelnou pohromu a na obnovu po havárii, a to včetně alespoň jedné zálohy všech uchovávaných informací mimo pracoviště.

V současné době neexistují specifické plány připravenosti na živelnou pohromu a na obnovu po havárii. Zálohovat digitální knihovnu mimo pracoviště nelze provádět z důvodu nedostatku financování (Jan Kolátor, 2017).

Digitální knihovna NTK požadavek normy **ISO 16363 v bodě 5.2.4. nesplňuje**.

3 Vyhodnocení auditu digitální knihovny NTK

3.1 Výstupy auditu digitální knihovny digitální knihovny NTK

Digitální knihovna NTK **celkově splňuje 50 z 93 požadavků normy ISO 16363**. Všechny poznatky získané auditem byly analyzovány a jednotlivým problémům, které v určité míře buď ohrožují digitální objekty, či snižují důvěryhodnost repositáře byly přiřazeny priority podle závažnosti.

Na základě těchto poznatků byly následovně stanoveny následující krátkodobé a dlouhodobé cíle v reakci na objevené nedostatky:

3.1.1 Krátkodobé cíle

Nejsnáze řešitelný problém jsou pravděpodobně **nedostatečné metadatové popisy některých objektů**. Z uživatelského hlediska jsou sice dostatečně popsány, nicméně absence některých technických metadat (zejména systémové číslo umožňující jednoznačnou identifikaci záznamu v katalogu) se může projevit jako podstatný problém např. při přebírání metadat skrze protokol OAI-PMH, používaného pro registr digitalizace nebo souborný katalog ČR. Existence systémového čísla v metadatech také umožňuje propojení katalogu NTK s dokumentem a tím jeho lepší zprostředkování cílové skupině a lepší správu objektů. Tento problém lze řešit programátorsky – vytvořením skriptu, který pomocí REST-API repositáře Fedora Commons, vygeneruje seznam všech záznamů v repositáři a exportuje jejich metadata. Tento seznam bude na základě křížového porovnání UUID identifikátorů Krameria a odkazů v katalogizačních záznamech vytvořit výstup rámcově 700 záznamů, u kterých systémové číslo chybí. Tyto záznamy bude potřeba upravit a vytvořit propojení mezi Krameriem a katalogem NTK. Aby se podobnému problému v budoucnosti předešlo, bude celý proces automatizován a nově přidané dokumenty budou do katalogu zapisovány ihned po jejich importu.

Dalším problémem, který lze vyřešit v krátkodobé rovině je **sepsání přesné dokumentace pracovních postupů**. Ty je možno z části čerpat z provozní dokumentace a není tudíž nutné vynakládat velké množství prostředků a času. Předpokládá se vytvoření dokumentace pro systém ProArc, Kramerius a LTP úložiště a zapojení samostatné komponenty RDFFlow systému ProArc, která by pomohla při sledování, standardizaci a průběhu digitalizace. Dále také k těmto informacím přidat popisy struktury SIP, AIP, DIP balíků a pravidla pro popis monografií a periodik. (Národní digitální knihovna, 2017).

3.1.2 Dlouhodobé cíle

Digitální knihovna NTK stojí před náročným a dlouhým procesem, má-li se vyrovnat certifikovaným repositářům, popř. se na samostatný proces certifikace sama připravit. V současné době se nepředpokládá, že by o certifikaci repositář v dohledné době usiloval. (Ondřej Koch, 2017)

Z hlediska priority je třeba **vytvořit a schválit strategii pro digitalizaci a dlouhodobou ochranu digitálních dokumentů**. Tato dokumentace poté umožní stanovovat další dlouhodobé cíle a formovat celkovou strategii provozování digitálního repositáře. První bude vytvořena strategie digitalizace, vzhledem k tomu že už existují její nepublikované základy a je potřeba ji pouze přepracovat tak, aby vyhovovala všem požadavkům ze strany digitální knihovny, tak vedení a bylo ji možné zveřejnit. Vývoj a definování strategie dlouhodobé ochrany bude probíhat společně s implementací úložiště dlouhodobé ochrany, aby byla zajištěna její přesná a vhodná formulace.

Samotné LTP úložiště a jeho procesy je pokládáno také za jeden z dlouhodobých cílů. Jako přechodné řešení je vlastní konfigurace, nicméně z dlouhodobého hlediska by bylo vhodné se zapojit, nebo alespoň sledovat vývoj projektu **ArcLib** ve spolupráci s Národní knihovnou ČR, Moravskou Zemskou knihovnou a Knihovnou Akademie Věd ČR na vývoji komplexního open-source řešení (Melichar, 2016) a následně jej **implementovat i v rámci instituce**.

V souvislosti s kontrolou udržitelnosti digitální knihovny a jejího rozvoje by měl personál **provádět minimálně 1x ročně opětovný audit** za pomoci normy použité pro účely této práce a případně doplňkově v rámci příbuzných nástrojů jako je Drambora nebo Nestor. Tyto audity budou zahrnuty pravidelných v hlášeních o stavu digitální knihovny a na jejich základě budou definovány další nové krátkodobé i dlouhodobé cíle. Digitální knihovna NTK by také měla **pokračovat ve spolupráci** s výše zmíněnými institucemi mimo jiné na vývoji systému Kramerius, udržovat a rozvíjet své technologické vybavení a tím držet krok s rozvojem nových technologií a sledovat aktuální rozvoj na poli standardizace a certifikace má-li zvyšovat svoji důvěryhodnost a udržitelnost na poli ochrany digitálních dokumentů.

4 Závěr

Principy a postupy na poli správy digitálních knihoven stále procházejí vývojem. Můžeme sledovat jejich rozčleňování do jednotlivých odvětví a profesí a jejich sdílená standardizace a normy na globální úrovni zlepšují kvalitu a odolnost digitálních dat. Normalizace a kodifikace je pro toto institučně a implementačně různorodé prostředí digitálních knihoven velmi důležitá (National Research Council, 2015). Otevřený archivační informační systém (ISO 16363:2012) se stal mezinárodně uznávaným rámcem definujícím základní funkce, povinnosti a strukturu repositářů a na jeho základě vznikly další normy pro certifikaci a auditní nástroje digitálních repositářů.

Instituce v České Republice sledují globální dění. Jednou z priorit Koncepce rozvoje knihoven na léta 2011–2015 bylo i trvalé uchovávání digitálních dokumentů. To umožnilo vytvoření Národní koncepce dlouhodobé ochrany digitálních dat pro knihovny (Ústřední knihovnická rada ČR, 2014). Vzniká také řada odborných publikací, konferencí, portálů a diskuzních skupin týkajících se problematiky digitálního kurátorství.

Audit objevil z části podobné výsledky jako u ostatních knihoven v minulých letech (Hruška, 2013; Hutař, 2008). Digitální knihovna NTK má oproti certifikovaným repositářům velké mezery zejména v rámci dokumentace a její transparentnosti, nejvíce u strategií digitalizace a dlouhodobé ochrany dokumentů. Strategie digitalizace byla v minulých letech podána jako návrh, nicméně ten nebyl schválen a byl vrácen k přepracování a doplnění. Tento dokument kromě samotného zvyšování důvěryhodnosti repositáře důležitý zejména z hlediska formulace způsobu, jakým digitální repositář získává dokumenty, poskytuje pro celý proces jasně definovaný rámec, čímž snižuje počet rizik, které by mohly ohrožovat následovnou archivaci digitálních dokumentů, celý proces standardizuje a urychluje. Dále také určuje, jakým způsobem jsou tyto dokumenty zpracovávány do dodavatelských balíčků, které mimo jejich transformace do archivních balíčků jsou i skrze pracovní postup zpřístupňovány uživatelům skrze webové rozhraní Krameria. Mimo jiné jasně definované pracovní postupy umožňují lepší zaučování nových či náhradních zaměstnanců v případě odchodu stávajících, což také snižuje riziko závažnějších rizik. Strategie dlouhodobé ochrany digitálních dokumentů je imperativní pro uchovávání, přístupnost a srozumitelnost digitálních dokumentů.

Dlouhodobá archivace není lehký úkol, vzhledem k tomu že časový horizont, ve kterém digitální dokumenty zastarávají, se neustále mění a může se, jak zkracovat, tak i prodlužovat (Cubr, 2010). Norma OAIS správně určuje tento horizont jako neurčitý

a tím pádem musí ochrana trvat tak dlouho, jak dlouho je nutné řešit dopady relevantních změn na archivované dokumenty (Consultative Committee, 2002). Tím že repositář takovouto strategii implementuje získává podobně jako u strategie digitalizace lepší důvěryhodnost ale opět i jasný rámec pro své pracovní postupy, transparentnost a hladší průběh řešení rizikových situací. Je důležité, aby oba dva tyto dokumenty byly v blízkém časovém horizontu vytvořeny a schváleny. Dalším, menším nedostatkem je definice pracovních postupů digitální knihovny. Fungování Krameria NTK (Digitální knihovna, 2017) a proces digitalizace (Digitalizační workflow NTK, 2017) je pro veřejnost popsán na webu, nicméně přímá dokumentace pro pracovníky je stručná, spíše informativního charakteru a mohla by se také vzhledem k její zastaralosti projevit jako problém při změně personálu. Samotné digitální objekty v digitální knihovně v jejich současném stavu se nejeví jako problémové, nicméně jako určité riziko se do budoucna může jevit právě starší část fondu digitální knihovny. Celkově lze dokumenty rozdělit do tří kategorií.

První kategorie, dokumenty z doby vzniku digitální knihovny jsou opatřeny jenom základními metadaty. Jedná se zejména o údaje o jméně, autorovi, vydání apod., nicméně nejsou opatřeny dostatečnými identifikátory pro jejich jednoznačné určení mimo digitální knihovnu a částí technických metadat.

Druhá skupina obsahuje metadata víceméně kompletní, ale neexistují, stejně jako první skupiny zálohy v kompletním balíku ale pouze skenů. Tento stav se jeví jako dlouhodobě neudržitelný a je potřeba dané objekty migrovat do struktury NDK-PSP balíčků a doplnit jejich popisnou informaci na úroveň třetí skupiny.

Třetí skupina obsahuje již nově vytvářené balíky s plným metadatovým popisem přebíraným z katalogu knihovny NTK. Právě digitální objekty spadající třetí kategorie se z hlediska auditu projeví jako kvalitní řešení, v jehož implementaci je třeba pokračovat.

Silné stránky digitální knihovny spočívají zejména v její technické a institucionální rovině. Systémy ProArc a Kramerius mají širokou uživatelskou základnu knihoven, které je využívají, jsou stabilní a po technické stránce velmi propracované. Jejich implementace v digitální knihovně NTK se osvědčila a jsou dlouhodobě používány bez jakýchkoliv obtíží.

Technické vybavení digitální knihovny je ve standardní kondici, dobře udržované a kvalitně monitorováno. V rámci instituce má digitální knihovna svoje pevné místo, je stabilně financována a je dobře podporována oddělením ICT infrastruktury. Přes veškeré nedostatky je vidět touha po jejím rozvoji, vytváření kvalitního zázemí a služeb pro uživatele a snaha o spolupráci s širokou škálou ostatních institucí na společných projektech. Z hlediska vývoje během posledního roku se zdá, že má nakročeno správným směrem.

5 Seznam použitých zdrojů:

1. BACA, Murtha. *Introduction to metadata*. 2nd ed. Los Angeles, CA: Getty Research Institute, 2008. ISBN 978-0-89236-896-9.
2. CAPLAN, Priscilla. 2008. The preservation of digital materials. *Library technology reports*. Feb/March 2008, vol. 44, issue 2, s. 4-38.
3. ČSN ISO 16363. *Systémy pro přenos dat a informací z kosmického prostoru – Audit a certifikace důvěryhodných digitálních úložišť*. Praha: Úřad pro technickou normalizaci, metrologii a státní zkušebnictví, 2014.
4. Consultative Committee for Space Data Systems. 2002. *Reference model for an open archival information system (OAIS)*: CCSDS Jan 2002. Blue book, Issue 1. Dostupné z: <https://siarchives.si.edu/sites/default/files/pdfs/650x0b1.PDF>
5. Consultative Committee for Space Data Systems. 2012. *Reference model for an open archival information system (OAIS)*: CCSDS 650.0-M-2 [Online]. Washington (DC, USA): CCSDS, May 2012. Magneta book. Dostupné z: <https://public.ccsds.org/Pubs/650x0m2.pdf>
6. CUBR, Ladislav, 2010. *Dlouhodobá ochrana digitálních dokumentů*. Praha: Národní knihovna České republiky. ISBN 978-8-07050-588-5.
7. DE BEER, Marietjie et al. Legal deposit of electronic books – a review of challenges faced by national libraries. *Library Hi Tech* [online]. 2016, vol. 34, no. 1, s. 87-103. ISSN 0737-8831.
8. Digitální knihovna. *Národní technická knihovna* [online]. Praha: NTK, 2017. Dostupné z: <https://www.techlib.cz/cs/2919-digitalni-knihovna>
9. Digitalizační workflow NTK. 2017. *Národní technická knihovna* [online]. Praha: NTK, 2017. Dostupné z: <https://www.techlib.cz/cs/2934-digitalizacni-workflow-ntk>
10. DRISCOLL, Erin. Presentation: Copyright and legal risks in digital preservation. In: *Digital preservation: where are we at, where are we going? Conference*. School of Information Studies, Charles Sturt University: National library of Australia, 2006.
11. GLADNEY, Henry M. *Preserving digital information*. New York: Springer, c2007. ISBN 978-3-54037-886-0.
12. HEDSTROM, Margaret. Digital preservation: a time bomb for digital libraries. *Computers and the Humanities*, 1997, 31.3: 189-202.
13. HRUŠKA, Zdeněk. Audit digitálních repozitářů. *Duha* [online]. 2013, roč. 27, č. 4. Dostupný z: <http://duha.mzk.cz/clanky/audit-digitalnich-repozitaru>. ISSN 1804-4255.

14. HUTAŘ, Jan, Andrea FOJTŮ a Eliška PAVLÁSKOVÁ. DRAMBORA – nástroj na interní audit digitálních repozitářů v nové online verzi a postřehy z provedených auditů. In: *INFORUM 2008: 14. konference o profesionálních informačních zdrojích: Praha, 28. – 30.5. 2008* [online]. 2008 Dostupné z: <http://www.inforum.cz/pdf/2008/hutar-jan-cze.pdf>
15. HUTAŘ, Jan a Marek MELICHAR. Dlouhodobá archivace digitálních dat - od teoretických úvah k praktické realizaci? *Knihovna: knihovnická revue*, 2015, 26(2), s. 58-68. ISSN 1801-3252. Dostupné také z: <http://knihovnarevue.nkp.cz/aktualni-cislo/knihovny-a-informace/dlouhodobaa-archivace-digitalnich-dat-2013-od-teoretickyh-uvah-k-prakticke-realizaci>
16. ISO 14721:2003. *Space data and information transfer systems -- Open archival information system -- Reference model*. Geneva: International Organization for Standardization, 2003.
17. ISO 14721:2012. *Space data and information transfer systems – Open archival information system (OAIS) – Reference model*. Geneva: International Organization for Standardization, 2012. 126 s. a ČSN ISO 14721. *Systémy pro přenos dat a informací z kosmického prostoru – Otevřený archivační informační systém – Referenční model*. Praha: Úřad pro technickou normalizaci, metrologii a státní zkušebnictví, 2014.
18. Knihovní řád a jeho přílohy. 2017. *Národní technická knihovna* [online]. Praha: NTK, 2017. Dostupné z: <https://www.techlib.cz/cs/2732-knihovni-rad>
19. KOCH, Ondřej. Ústní sdělení. Praha: Národní technická knihovna, 13. 2. 2017.
20. KOLÁTOR, Jan. Ústní sdělení. Praha: Národní technická knihovna, 13. 2. 2017
21. Komplexní validátor NDK. 2017. *Národní digitální knihovna* [online]. Praha: NK ČR, 2017. Dostupné z: <http://www.ndk.cz/archivace/komplexni-validator>
22. Kramerius Wiki. *GitHub.com* [online]. Praha: INCAD, s.r.o., 2017. Dostupné z: <https://github.com/ceskaexpedice/kramerius/wiki>
23. *Kramerius NTK* [online]. Praha: NTK, 2017. Dostupné z: <http://k4.techlib.cz/search/#intro3>
24. KVAŠOVÁ, Zuzana; VAŠEK, Zdeněk. Národní koncepce dlouhodobé ochrany digitálních dat v knihovnách. *Bulletin SKIP* [online]. 2017, roč. 26, č. ISSN 1213-5828. Dostupné z: <https://bulletinskip.skipcr.cz/node/85>
25. Marc standards. *Library of Congress* [online]. Washington DC: Library of Congress, 2017. Dostupné z: www.loc.gov

26. MELICHAR, Marek, Jan HUTAŘ, Michal RŮŽIČKA, Zdeněk HRUŠKA, Zdeněk VAŠEK, Miroslav BARTOŠEK a Martin LHOTÁK. *Projekt ArcLib – budování systému pro dlouhodobou archivaci digitálních dat v českých knihovnách*. IT lib, Bratislava: Centrum vedecko-technických informací SR, 2016, roč. 20, č. 2, s. 13-17. ISSN 1335-793X.
27. MINISTERSTVO KULTURY. *Koncepce rozvoje knihoven ČR na léta 2011–2015 včetně internetizace knihoven* [online]. 2012. Dostupné z: http://www.mkcr.cz/assets/literatura-a-knihovny/Koncepce_rozvoje_knihoven_2011-2015.pdf
28. NÁRODNÍ TECHNICKÁ KNIHOVNA. 2016. Zápis č. 10/2016 z kolegia ředitele NTK: interní dokument. Praha, 2016.
29. NÁRODNÍ TECHNICKÁ KNIHOVNA. 2016. Katalogizační politika NTK: interní dokument. Praha, 2016.
30. Národní koncepce dlouhodobé ochrany digitálních dat v knihovnách. *Ústřední knihovnická rada ČR* [online]. 2016. Dostupné z: <http://files.ukr.knihovna.cz/200000240-eb067ec003/KoncepceUchovDigit2016.docx>
31. Národní digitální knihovna. *Standardy digitalizace NDK*. [online]. Praha: Národní knihovna, 2017. Dostupné z: <http://www.ndk.cz/standardy-digitalizace/metadata>
32. NÁRODNÍ ULOŽIŠTĚ ŠEDÉ LITERATURY. *Audit Národního úložiště šedé literatury (NUŠL) v NTK s použitím nástroje DRAMBORA: Třetí audit 2011*. Praha, 2011. Dostupné z: http://repozitar.techlib.cz/record/389/files/idr-389_1.pdf
33. POLČÁK, Radim. *Digitalizace knihovního fondu NTK*. Masarykova univerzita, Právnická fakulta, 2009. Dostupné také z: <http://repozitar.techlib.cz/record/78>
34. National Research Council of the National Academies. *Preparing the workforce for digital curation*. Washington (DC): National Academies Press, 2015. ISBN 030-9-29694-3.
35. Preserving digital information: report of the Task Force on Archiving of Digital Information. Washington, D.C.: Commission on Preservation and Access, 1996. ISBN 188-7-33450-5.
36. RIEGER, Oya Y. Preserving and Emulating Digital Art Objects. In: *Articles, Preprints, and Presentations by CUL Staff* [online]. Cornell: Cornell University Library, 2015, s. 202. Dostupné z: <http://hdl.handle.net/1813/41368>
37. ROBERTSON, Steven B. *Digital Rosetta Stone: a conceptual model for maintaining long-term access to digital documents*. AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH, 1996.

38. ŘIHÁK, Jakub. *Automatizace digitalizačního workflow NTK*. Praha, 2014. Dostupné také z: <https://www.vse.cz/vskp/eid/41374>. Diplomová práce. Vysoká škola ekonomická v Praze. Vedoucí práce Kosek, Jiří.
39. RYLICH, Jan. TransISTor 2010 aneb Ochrana elektronických dokumentů a audiovizuálních děl. *Ikaros* [online]. 2010, ročník 14, číslo 7. urn:nbn:cz:ik-13516. ISSN 1212-5075. Dostupné z: <http://ikaros.cz/node/13516>
40. Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report [online]. RLG: Mountain View (CA), 2002. 70 s. Přístup z: <https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>
41. Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist [online]. Dublin (OH); Chicago (IL): OCLC, CRL, 2007. 94 s. . Dostupné z: http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf
42. Zákon o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon). In: *121*. Praha: Parlament ČR, 2000, ročník 2000, částka 36, číslo 121.

6 Seznam obrázků:

Obrázek 1 - Referenční model OAIS (Rylich, 2010)	5
Obrázek 2 - Datový tok a prostředí OAIS	7

7 Seznam grafů

Graf 1 - Počet uživatelů digitální knihovny NTK 2013–2016	25
Graf 2 - Návštěvnost digitální knihovny NTK 2013–2016	25
Graf 3 - Souhrn návštěvnosti digitální knihovny NTK 2013–2016	25

8 Seznam zkratek:

AIP – Archival information package

API – Application Programming Interface

CCSDS – Consultative Committee for Space Data Systems

ČSN – Český normalizační úřad

DIAS – Digital Information Archiving Solution

DIP – Dissemination information package

DOI (Digital object identifier)

Drambora – Digital Repository Audit Method Based on Risk Assessment

ICT – Information and Communication Technologies

ISO – International Organization of Standardization

LTP – long term preservation

METS – Metadata Encoding and Transmission Standard

MODS – Metadata Object Description Schema

NDK – Národní digitální knihovna

Nestor – Network of Expertise in Long-Term Storage of Digital Resources

NTK – Národní technická knihovna

OAI-PMH – Open Archives Initiative - Protocol for Metadata harvesting

OAIS – Open Archival Information System

OCLC – Online Computer Library Center

PLATTER – Planning Tool for Trusted Electronic Repositories

PREMIS – PREservation Metadata: Implementation Strategies

ProArc – Production and Archival system

REST – Representational State Transfer

RLG – Research Library Group

SDB – Safety Deposit Box

SIP – Submission Information Package

TRAC – Trustworthy Repositories Audit & Certification

URN: NBN - Uniform resource name : national bibliography number

UUID - universally unique identifier

9 Seznam příloh

Příloha 1 – Smlouva o digitalizaci dokumentů NTK

Příloha 2 – Výstup Google Analytics digitální knihovny NTK 2013-2016