NÁRODNÍ KONCEPCE DLOUHODOBÉ OCHRANY DIGITÁLNÍCH DAT V KNIHOVNÁCH

5. 12. 2016

Jako výstup Priority č. 2 Koncepce rozvoje knihoven ČR na léta 2011 – 2015 (Vytvořit legislativní, organizační a technické předpoklady trvalého uchování a zpřístupnění publikovaných digitálních a digitalizovaných dokumentů jako důležité složky kulturního a vědeckého dědictví) zpracovali pro Ústřední knihovnickou radu v rámci implementace Koncepce rozvoje knihoven ČR na léta 2017–2020 a s výhledem do roku 2025 Jan Hutař (Archives New Zealand), Marek Melichar a Tomáš Gec (oba Univerzita Karlova).

OBSAH

1 Uvedení do problému ... 4
1.1 Kulturní digitální data v knihovnách ... 4
1.2 Současná legislativa ... 4
1.3 Dlouhodobá ochrana digitálních dat jako oblast výzkumu a praxe ... 5
1.4 Současný stav v ČR ... 6
1.5 Produkční standard NDK ... 7
1.6 Národní koordinace LTP v jiných zemích ... 9
1.7 Česká situace jako východisko pro formulaci strategie ... 10
2 Definice cílů a principů národní strategie ... 12
3 Ideální stav za pět let v oblasti dlouhodobé ochrany ... 14
3.1 Jádro systému pro dlouhodobou ochranu knihovních dat ... 14
3.2 Další certifikované systémy ... 15
3.3 Otevřená řešení ... 15
3.4 Born-digital data ... 16
3.5 Vzdělávání, výzkum a certifikace ... 16
3.6 Sdílení a financování ... 17
4 Konkrétní doporučená opatření ... 18
4.1 Plán na zajištění dlouhodobé ochrany digitálního kulturního a vědeckého dědictví v ČR ... 18
4.2 Garantovaná datová centra pro zajištění bitové ochrany digitálních dat ... 18
4.3 Garantovaný systém pro zajištění logické ochrany digitálních dat ... 18
4.4 Postupné budování sítě certifikovaných úložišť v jednotlivých institucích ... 19
4.5 Podpora otevřených řešení pro zajištění logické ochrany digitálních dat ... 19
4.6 Zřízení Metodického centra pro dlouhodobou ochranu digitálních dat ... 20
4.7 Krajské knihovny jako producenti a správci digitálních dat ... 20
4.8 Malé instituce jako producenti dat ... 21
4.9 Finanční mechanismy ... 21
4.10 Organizační a systémové změny ... 22
4.11. Konsolidace dříve pořízených dat a interoperabilita ... 22
5 Doporučení v oblasti zajištění kvality, standardizace a certifikace ... 23
5.1 Zajištění kvality; audit a certifikace ... 23
5.2 Doporučené postupy ... 24
6 Postup realizace koncepce ... 25

6.1 Optimální postup realizace koncepce 25
První fáze (2016) 25
Druhá fáze (2017–2018) 25
Třetí fáze (2019–2020) 25
6.2 Minimální očekávaný stav v roce 2020 26
7 Seznam použitých zkratek 27
8 Seznam citovaných norem 29

1 Uvedení do problému

1.1 Digitální data v knihovnách

Knihovny se musí dlouhodobou ochranou[1] digitálních dat zabývat především proto, aby v digitální době přežily jako klíčové informační instituce. Jejich uživatelé dnes přicházejí především po síti a očekávají plnohodnotné služby v digitální a online podobě. Knihovny se jim musí přizpůsobit a poskytovat online jak aktuální informační zdroje (článkové databáze, webové archivy), tak dokumenty, které vznikají výhradně v digitální podobě (born-digital) a nahrazují tak dokumenty klasické. Knihovny musí tyto dokumenty dostupné výhradně v digitální podobě získávat, ochraňovat a zpřístupňovat. Vedle toho uživatelé automaticky předpokládají, že knihovny online (tedy v digitalizované podobě) nabídnou i dokumenty, které jsou součástí jejich fondů již z minulosti. Kromě výše uvedeného musí knihovny (např. vysokoškolské) stále častěji přijímat a zpřístupňovat i nové typy objektů (vědecká data, audiovizuální dokumenty). Knihovny byly po staletí zvyklé pracovat s fyzickými dokumenty. V digitálním světě takto dlouhou zkušenost zatím nemají. Pokud ale chtějí jako instituce přežít, musí se digitálnímu světu přizpůsobit. Motivace zabývat se dlouhodobou ochranou digitálních dat ale přichází i zvenčí. Česká republika by jako členská země EU měla následovat evropská doporučení. Dlouhodobá archivace kulturního dědictví v digitální podobě je jednou z klíčových aktivit Digitální agendy pro Evropu[2], která pokrývá jak oblast digitalizace, tak dlouhodobou ochranu a zpřístupnění.

1.2 Současná legislativa

Povinnost ukládat a ochraňovat dokumenty uložené v knihovnách pro budoucí generace pramení z knihovního zákona (Zákon o knihovnách a podmínkách provozování veřejných knihovnických a informačních služeb č. 257/2001 Sb., ve znění pozdějších změn). Tento zákon se vztahuje na knihovny, které jsou zapsány v evidenci knihoven MK. V § 18 se mluví o ochraně knihovního fondu. Knihovní fond a knihovní dokumenty musejí být ochráněny před poškozením a zcizením; uloženy ve vhodných podmínkách a provozovatel knihovny musí zajistit převod dokumentu na jiný druh nosiče, pokud je to třeba k jejich uchování.

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[3].

---

[1] Jsou též používány termíny douhodobá archivace či dlouhodobé uchovávání.
[2] EVROPSKÁ KOMISE. *Digitisation & Digital Preservation* [online]. [Brussels]: European Commission, Last updated on 19/07/2016 – 16:30 [cit. 2016-12-05]. Dostupné z: https://ec.europa.eu/digital-agenda/en/digitisation-digital-preservation
[3] Knihovním dokumentem [je] informační pramen evidovaný jako samostatná jednotka knihovního fondu knihovny. Knihovním fondem [je] organizovaný, soustavně doplňovaný, zpracovávaný, ochraňovaný a uchovávaný soubor knihovních dokumentů.

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.[4]

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).

1.3 Dlouhodobá ochrana digitálních dat jako oblast výzkumu a praxe

Pro trvalé uchování digitálního kulturního dědictví je třeba zavést organizační procesy a (technické) informační systémy, které umožní vybudování systémů pro dlouhodobou ochranu digitálních dat (LTP systémů), tj. systémů, které budou schopny zajistit udržení a použitelnost digitálních dat ve stejném horizontu, na jaký jsme zvyklí u fyzických dokumentů – po dobu několika set let.

O dlouhodobé ochraně digitálních dat se začalo hovořit již v polovině 80. let 20. století, kdy si někteří archiváři začali uvědomovat, že s uchováním digitálních dat to nebude tak jednoduché: že ochranu digitálních dat je třeba dobře plánovat a je třeba počítat s neustálými zásahy a přesuny dat, protože digitální technologie mají krátkou životnost.[5] Velmi důležitou roli sehrála americká zpráva *Preserving Digital Information*, která vznikala od roku 1994 a byla vydána v roce 1996[6]. Jedná se o jeden z prvních dokumentů, v němž byla popsána dlouhodobá ochrana digitálních dat v dnešním smyslu, a to včetně problematiky zastarávání softwaru i hardwaru, možností řešení (migrace), otázek integrity digitálních dokumentů, jejich provenience a zachycení kontextu atp. V závěru zprávy autoři konstatovali, že v době jejího vydání neexistovala žádná vhodná metoda na zajištění dlouhodobé ochrany digitálních informací a společnost ani její instituce nejsou na tuto problematiku připraveny. Reakce odborníků na oblast archivnictví na sebe nedala dlouho čekat. V roce 1999 vznikl první návrh referenčního rámce OAIS (*Open Archival Information System*), pozdější normy ISO 14721[7]. Ta je dodnes základním orientačním bodem pro každého, kdo se dlouhodobou ochranou vážně zabývá. Další vývoj pak vedl ke standardizaci pojetí dlouhodobého důvěryhodného digitálního archivu (viz ISO 16363:2012 včetně prováděcí normy ISO 16919).

Dlouhodobá ochrana digitálních dat (anglicky nejčastěji označovaná jako *digital preservation* nebo šířeji jako

---

[4] Archiválie podle bodu f) stejného paragrafu je „takový dokument, který byl vzhledem k době vzniku, obsahu, původu, vnějším znakům a trvalé hodnotě dané politickým, hospodářským, právním, historickým, kulturním, vědeckým nebo informačním významem vybrán ve veřejném zájmu k trvalému uchování a byl vzat do evidence archiválií, …".

[5] Viz např. FOX, Lisa L. Archival Preservation in the Age of Technology. *Provenance, Journal of the Society of Georgia Archivists* [online]. 1985, vol. 3, issue 1, article 4 [cit. 2016-12-05]. Dostupné z: http://digitalcommons.kennesaw.edu/provenance/vol3/iss1/4/

6 GARRETT, John a Donald WATERS. *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* [online]. The Commission on Preservation and Access and The Research Libraries Group, 1996 [cit. 2016-12-05]. Dostupné z: http://www.clir.org/pubs/reports/pub63watersgarrett.pdf

[7] V textu jsou označení norem uváděna ve zkráceném tvaru bez roku vydání, podrobné údaje jsou k dispozici v seznamu norem v závěru dokumentu.

digital curation) je oblastí výzkumu[8] a v posledních letech také každodenní praxe v řadě institucí.

V digitálním světě, stejně jako ve světě fyzických dokumentů, není možné uchovávat všechny dokumenty. Knihovny musí jasně stanovit, které objekty mají být předmětem dlouhodobé ochrany a musí rozlišovat, která data skutečně má smysl ukládat dlouhodobě a s vynaložením maximální péče a která lze ukládat bez zaručení dlouhodobé udržitelnosti.[9]

Nutným předpokladem a prvním krokem k dlouhodobé ochraně digitálních dokumentů a vytvoření dlouhodobého důvěryhodného digitálního archivu je zajištění ochrany na úrovni bitů. Knihovny, které mají vlastní datová úložiště a datová centra, vědí, že to znamená zajištění neměnnosti dat, fyzickou ochranu, zálohování, sledování a obnovu technologií a vícenásobné ukládání dat, tj. ukládání na více médiích ve více lokalitách.

Druhým krokem je zajištění logické ochrany, jejímž cílem je zajištění použitelnosti, srozumitelnosti, zobrazitelnosti archivovaných digitálních dokumentů v jakémkoliv okamžiku v budoucnu bez ohledu na změny technologií. Toho je dosaženo vytvářením a ukládáním technických, ochranných, administrativních metadat po celou dobu existence dokumentu. Tato metadata jej popisují, udržují kontext a lze je využít např. při převodu dokumentu do jiného formátu apod. Logická ochrana digitálních dat předpokládá nejen technické systémy, ale hlavně především nastavení procesů podle požadavků norem ISO 14721:2012 a ISO 16363:2012.[10]

Vedle standardizace organizačních procesů je také nutné mít zajištěné financování, odborné pracovníky, odpovídající prostory aj. To jsou vše příklady požadavků z normy ISO 16363 na to, aby mohlo být digitální úložiště certifikováno jako důvěryhodné.

1.4 Současný stav v ČR

V České republice se knihovny dlouhou dobu zaměřovaly především na tvorbu nebo sklízení digitálního obsahu – projekty financované z veřejných rozpočtů podporovaly vytváření digitálních dat (digitalizace, archivace webu), přičemž dlouhodobá ochrana těchto data byla podceňována. V současnosti je v této oblasti situace výrazně fragmentovaná, projekty mají různé cíle, nejčastěji zpřístupnit data v digitalizované podobě, vzácně ochránit digitální data pro budoucnost. Z velkých digitalizačních projektů probíhajících v ČR (NDK[11],

[8] Tato oblast je již od konce 90. let 20. století podporována z veřejných rozpočtů. V EU byly z 5. rámcového programu (FP5; 1998–2001) financovány projekty ERPANET, DigiCULT, PRESTO, ECHO a MINERVA, z 6. rámcového programu (FP6; 2002–2006) projekty PRESTOSPACE, DELOS, BRICKS, CASPAR, DPE, PLANETS a PRESTOSPACE a konečně ze 7. rámcového programu (FP7; 2007–2013) projekty PROTAGE, DL.ORG, LIWA, SHAMAN, KEEP, PRESTOPRIME, SCAPE, ARCOMEM a TIMBUS. V USA se jednalo např. o projekt NDIIPP, na Novém Zélandě o projekt NDHA. Byla realizována i řada dalších národních projektů.

[9] Dočasně ukládaná data musí být možné nahradit (přeskenováním, konverzí z masterů atd.).

[10] Je tedy důležité mít procesní dokumentaci, strategii, plány pro různé situace, dokumentaci pravidelně revidovat apod.

[11] NÁRODNÍ KNIHOVNA ČESKÉ REPUBLIKY. *NDK: Národní digitální knihovna* [online]. Praha: Národní knihovna ČR [cit. 2016-12-05]. Dostupné z: http://www.ndk.cz/

Google Books[12], Manuscriptorium[13]) dlouhodobou ochranu přímo řeší pouze NDK. Dále existují menší digitalizační aktivity (program VISK 7[14], krajské digitalizace, Norské fondy), které o dlouhodobé ochraně vytvořených dat alespoň uvažují. Aktivní jsou na tomto poli NK ČR a MZK. KNAV spolu s dalšími institucemi vyvíjí v projektu NAKI II otevřený (*open source*) systém ArcLib, který by měl být volně dostupným řešení v oblasti dlouhodobé archivace pro ostatní knihovny. Některé další instituce se začínají dlouhodobou archivací zabývat (NTK, VŠCHT), určitou roli v této oblasti sehraje také CESNET.

Velkým úspěchem je prosazení standardní podoby výstupu digitalizace (tzv.standard NDK) na národní úrovni. Data a metadata vytvořená podle tohoto standardu jsou do jisté míry na dlouhodobou ochranu připravena, i když nebudou uložena v odpovídajícím LTP systému.[15]

Množství projektů, ve kterých digitální data vznikají, vede k tomu, že různé projekty používají různé metody zajištění kvality digitalizace a ukládání. Hlavním problémem je neexistence jasně stanoveného požadavku zajistit ve všech projektech odpovídající trvalé uložení klíčových digitálních dat. . Není jasně stanoveno, která data je nutné dlouhodobě ochraňovat (z důvodu možnosti pořízení dostatečně kvalitních kopií těchto dat), a jakým způsobem má být ochrana zajištěna, tj. jaké jsou požadavky na systémy, ve kterých mají tato data být ukládána

Data, která tvoří jádro kulturního dědictví (např. born-digital data; data z archivace webu; určité typy digitalizovaných dat), je třeba uchovávat v certifikovaných dlouhodobých digitálních úložištích. Všechna tato digitální data vytvářená a shromažďovaná v projektech financovaných z veřejných rozpočtů by tedy měla splňovat stanovené požadavky nutné k zajištění kvality jejich dlouhodobé ochrany. V současné době jsou v různých projektech uplatňovány různé představy o kvalitě, způsobu uložení dat a metadat, které ne vždy plně tyto požadavky plní.

Neznamená to volání po jednotném systému a jediném úložišti, naopak různorodost přístupů a systémů sníží závislost na jednom dodavateli a riziko jednoho typu řešení. Jedná se spíše o koordinaci na úrovni zajištění kvality (certifikace úložiště, standardy pro data, procesy), financování a metodik, tedy jasný návod pro rozlišení těch dat, kterým je třeba věnovat maximální péči.

1.5 Produkční standard NDK

Prosazení již zmíněného standardu NDK jako národního standardu pro digitalizaci v knihovnách považujeme za silnou stránku současného stavu. Formulace standardu NDK navazovala na zkušenosti s předchozí generací národního standardu postaveného na starších formátech DTD. Podle standardu DTD byly v NK ČR

---

[12] NÁRODNÍ KNIHOVNA ČESKÉ REPUBLIKY. *Národní kulturní bohatství zachovají nové technologie: Statisíce knih a periodik online* [online]. Praha: Národní knihovna ČR, 22. 11. 2011 [cit. 2016-12-05]. Tisková zpráva. Dostupné z: http://www.nkp.cz/soubory/ostatni/tz-ndk-google.pdf/at_download/file

[13] NÁRODNÍ KNIHOVNA ČESKÉ REPUBLIKY. [2016]. *Manuscriptorium: Digital Library of Written Cultural Heritage* [online]. [Praha]: Národní knihovna ČR [cit. 2016-12-05]. Dostupné z: http://www.manuscriptorium.com/

[14] Kopii dat z podprogramu VISK 7 archivuje NK ČR a do budoucna by měly být zahrnuty do LTP řešení NK ČR.

[15] Metadata již z procesu digitalizace obsahují základní technická a administrativní metadata, záznamy o procesech digitalizace apod. Tyto údaje pak mohou posloužit pro procesy dlouhodobé ochrany nebo jako vstup do LTP systému.

a v řadě dalších institucí dlouhá léta vytvářeny archivní balíčky novodobých digitalizovaných publikací. Standard postavený na DTD pro periodika a pro monografie byl závazný pro digitalizaci v rámci projektů z program VISK 7, využívala ho ale mj. také KNAV. Dodavatelské firmy upravovaly své nástroje tak, aby byly schopné metadata a data v určené podobě vytvářet. Data v těchto strukturách jsou často velmi cenná, jedná se především o digitální kopie nejohroženějších periodik.

V roce 2013 byl standard DTD jak v NK ČR, tak v rámci programu VISK 7 nahrazen standardem NDK.[16] Ten je založen na v současnosti obvyklých a doporučovaných formátech metadat (METS, PREMIS, MODS, Dublin Core, MIX a ALTO XML). Všechny jmenované standardy jsou de facto standardy, vyvíjené a podporované Kongresovou knihovnou USA. Jsou bez výjimky využívány ve všech projektech dlouhodobé archivace v knihovnách. Metadatová specifikace vznikala na základě podobných projektů a ve spolupráci s národními knihovnami Nizozemska, Norska a Finska a s Kongresovou knihovnou USA.

Prosazení standardu NDK na národní úrovni a uplatnění v projektech z programu VISK 7, při digitalizaci v krajích a jinde považujeme za velký úspěch. Ať se jedná o jakýkoliv projekt, základní úroveň ochrany je použitím tohoto standardu zajištěna, stejně jako tolik žádaná konzistence dat zajišťující interoperabilitu mezi projekty a jejich úložišti. Standard NDK nepokrývá pouze metadata, stejná pozornost byla věnována specifikacím obrazových dat a vytváření profilu JPEG 2000 a souborů OCR ve schématu METS ALTO. Standard NDK odpovídá současné nejlepší praxi v paměťových institucích ve světě. Ukládání jednotlivých naskenovaných stran jako digitalizačních masterů ve formátu JPEG 2000 představuje posun k formátu, který poskytuje lepší vizuální možnosti reprezentace obsahu než starší JPEG nebo TIFF. Je také prokazatelně odolnější vůči bitovým ztrátám než jiné formáty (PNG, JPEG) a ve srovnání s formátem TIFF při zachování stejné kvality výrazně (několikanásobně) šetří prostor. Ačkoli se nejedná o hojně používaný formát, jeho obliba jako archivního formátu pro digitalizovaná data roste a je používán v řadě knihoven a archivů (např. v Britské knihovně, Kongresové knihovně v USA, Britském národním archivu, národních knihovnách Nizozemska a Norska či ve Wellcome Library). Mnohé instituce z formátu TIFF do formátu JPEG 2000 převádějí svá archivní data, a to často z důvodu úspory prostoru na datovém úložišti a větší flexibility formátu JPEG 2000. Z pohledu vhodnosti pro dlouhodobou ochranu jsou formáty TIFF i JPEG 2000 považovány za rovnocenné.

Služba ČIDLO (systém pravidel a nástroj resolver) běží v ostrém provozu již od roku 2012 a přiděluje identifikátory URN:NBN aktuálně digitalizovaným monografiím a článkům periodik (v rámci projektu Vytvoření Národní digitální knihovny i mimo něj). Registrátory systému ČIDLO (institucemi, kterým je přidělován identifikátor URN:NBN) jsou v současnosti nejen NK ČR a MZK, ale téměř sto dalších knihoven České republiky. 5. června 2015 byla MK schválena certifikovaná metodika „Metodika pro přidělování a správu životního cyklu unikátních perzistentních identifikátorů digitálních dokumentů podle standardu URN:NBN“[17], která popisuje systém ČIDLO a pravidla jeho užití a byla vypracována NK ČR. Přidělování identifikátorů URN:NBN v knihovnách, které provádějí digitalizaci svých fondů, je nyní též povinným prvkem

[16] V letech 2011 a 2012 bylo v programu VISK 7 povoleno používat oba standardy.

[17] VAŠEK, Zdeněk et al. *Metodika pro přidělování a správu životního cyklu unikátních perzistentních identifikátorů digitálních dokumentů podle standardu URN:NBN* [online]. Praha: Národní knihovna ČR, [2016] [cit. 2016-12-05]. Dostupné z: http://www.ndk.cz/archivace/Metodika_URNNBN_final_2.1a.pdf

standardu NDK. Systém ČIDLO a certifikovaná metodika pro přidělování identifikátorů URN:NBN rovněž výrazně napomáhají prosazování cílů logické ochrany digitálních dokumentů v českém knihovnickém prostředí.

Standard NDK v současnosti pokrývá skenovaná periodika a monografie. Na doplňování a aktualizaci standardu je nutné trvale pracovat (přizpůsobení aktuální verzi standardu PREMIS atd.). Pro další udržování a zlepšování standardu vytvořila NK ČR tzv. Formátový výbor. Ten by měl stanovit požadavky na vytvoření dalších standardů pro vstupní (SIP) balíčky pro další typy digitalizovaných dat a digitálních dat. Cílem této pracovní skupiny je vytvořit komunikační platformu pro výměnu zkušeností uživatelů standardu NDK a formulování budoucích cílů v oblasti standardizace digitálních dokumentů[18].

Jako základ by měl být (i pro archivní účely) využíván standard METS, který obsahuje metadata ve schématech PREMIS, MODS a DC. Archivní balíček by oproti balíčku z digitalizace měl být obohacen o technická a administrativní metadata potřebná pro efektivní ochranu daného typu dokumentu.

1.6 Národní koordinace LTP v jiných zemích

V různých zemích ke koordinaci aktivit spojených s dlouhodobou archivací na národní úrovni přistupují velmi odlišně. EU v rámci svých aktivit, které jinak doposud směřovaly spíše k podpoře výzkumu a vývoje v oblasti dlouhodobé archivace, realizovala šetření v rámci projektu ENUMERATE, kde byly mj. pokládány otázky týkající se dlouhodobé archivace. V podstatě to jsou jedna z mála tzv. tvrdých dat o praxi v jednotlivých institucích v EU a ukazují, že evropské paměťové instituce sice aktivně digitalizují a pořizují digitální data, ovšem dlouhodobou archivací se skutečně zabývá nejvýše jedna čtvrtina z nich[19].

V různých zemích probíhá koordinace dlouhodobé archivace na národní úrovni různě. Někde se snaží koordinovat aktivity v archivech i knihovnách (Kanada, Nový Zéland a částečně také Nizozemsko), obvykle v rámci celonárodního projektu přechodu státu na digitální agendu. Jinde knihovny více spolupracují s akademickou sférou a společné projekty vedou buď ke sdílení znalostí a podpoře výzkumu (Spojené království[20]; Nízozemsko či Polsko). V některých zemích jsou pro dlouhodobou archivaci v paměťových institucích klíčové aktivity některé dominantní instituce (Dánsko, Norsko, Francie, Finsko). V jiných zemích je zájem o dlouhodobou archivaci diktován požadavky a projekty shora (Slovenský program OPIS a v rámci něj projekt CDA), jiné země nebo instituce budují systémy zdola (Sasko, resp. celé Německo, dále Polsko). Inspirativním příkladem pro Českou republiku mohou být Polsko nebo některé německé země, v nichž k budování služeb pro paměťové instituce využívají kapacity a dovednosti akademické sféry a akademická výpočetní centra.

---

[18] Např. vytvoření standardu pro vstupní balíček hudebnin, zvukových dokumentů, e-born dokumentů, archivaci webu nebo šedou literaturu.

[19] *ENUMERATE Focus Group Meeting on Digital Preservation: Monitoring and Measuring Digital Preservation practices in the EU cultural heritage domain* [online]. The Hague, 17 January 2013 [cit. 2016-12-05]. Dostupné z: http://www.enumerate.eu/fileadmin/ENUMERATE/meetings/ENUMERATE-FocusGroup-on-Digital-Preservation-Monitoring_TheHague_20130117_v01.pdf

[20] Jedná se např. o organizaci Digital Preservation Coalition.

Zvláštní postavení mají Německo, Spojené království a USA. V těchto zemích se dlouhodobé ochraně digitálních dat věnuje více než jedna instituce a znalostních center je tak více. V ostatních zemích se většinou jedná o národní instituci (knihovnu, archiv nebo jinou ústřední instituci). V Německu, USA a Spojeném království jsou vedoucími institucemi jak národní knihovny, tak národní archivy. Celou síť dále doplňují aktivní univerzity a datová centra (např. pro vědecké informace). Lze říci, že koordinace v těchto zemích neprobíhá.

1.7 Česká situace jako východisko pro formulaci strategie

- **Existence programu VISK** – díky existenci programu VISK (konkrétně VISK 7) má v České republice řada institucí zkušenosti s digitalizací (byť často externí, dodavatelskou) novodobých i historických fondů. Také další aktivity a finanční zdroje byly v České republice věnovány obvykle na pořizování digitálních dat (Norské fondy) nebo jejich zpřístupňování (Digitální knihovna Kramerius). Tyto finanční zdroje po institucích vždy vyžadují dodržení standardů pro tvorbu metadat i formáty digitalizace. Projekt digitálního úložiště pro muzea a galerie CITeM byl zastaven.
- **Současná spolupráce a možní partneři knihoven** – v České republice existuje silná infrastruktura akademické sítě a jsou budována akademická úložiště (CESNET). Akademická sféra se věnuje především zpřístupňování výsledků vědecké činnosti a kvalifikačních prací s využitím otevřených technologií (repozitáře DSpace, Invenio apod.). Bohužel například aktivity spojené s archivací vědeckých dat (především tzv. malých vědeckých dat z oblasti humanitních a společenských věd) – jsou běžné zatím pouze v zahraničí, u nás nikoliv. Vybudování datového archivu podobného např. britskému UK Data Archive, nizozemskému DANS nebo německému GESIS (Lipsko) není ze strany věcně příslušného ministerstva zatím nijak podporováno. Spolupráce mezi paměťovými institucemi (knihovny, muzea a galerie, a archivy) a akademickou sférou v oblasti dlouhodobé archivace je stále v plenkách. Těžko najdeme společné projekty výpočetního centra vysoké školy a krajského muzea.
- **Archivace webu** – Česká republika má jednu z nejstarších tradic archivace internetu v Evropě. Objemy dat z archivace webu jsou enormní a financování ochrany těchto dat by mělo dostat jasný právní rámec (statut Webarchivu, jasné přidělení odpovědnosti, závazek státu financovat ochranu těchto dat) s tím, že Webarchiv by měl směřovat k prokázání kvality své činnosti (dodržení norem ISO 14 873 a ISO 28500).
- **Stáří dat** – české kulturní instituce mají ve svých sbírkách digitalizovaná data od roku 1996 a data z archivace webu od roku 2001. Tato data jsou z technického hlediska velmi stará, dlouhodobá ochrana digitálních dat přesto donedávna neměla pevné místo v žádné z českých institucí.
- **Zdroje digitalizovaných dat** – vedle menších projektů financovaných z rozpočtů MK ČR v rámci programu VISK 7 je zde rozhodující role NK ČR a MZK a jejich projektu NDK, který usiluje u vybudování dlouhodobého důvěryhodného úložiště v souladu s ISO 16363.
- **Role MK** – role MK je pro paměťové instituce a jejich snahy zabývat se dlouhodobou archivací zcela klíčová – MK je často jejich zřizovatelem nebo vytváří finanční nástroje a pravidla, které by mohly

jejich činnost v oblasti dlouhodobé archivace podpořit. Financování (mj. program VISK 7) bylo vždy spojeno s tlakem na dodržení standardů a do budoucna by mělo být provázeno požadavkem zajištění dlouhodobé archivace.

- **Aktivity institucí** – ohnisek zájmu o dlouhodobou ochranu digitálních dat je nyní více (KNAV, NFA, NA a některé další archivy, MZK, CESNET, MU, FAMU, NDK, NTK, VŠCHT, UK) a tento zájem je třeba kultivovat a podpořit. Kvůli neexistenci podpory z ministerstev či vlády jde o zájem zdola. Také díky tomu se důvody tohoto zájmu (někdy i podstatně) liší. Konkrétní instituce mají prostřednictvím některých svých zaměstnanců velmi dobré znalosti problematiky dlouhodobé ochrany digitálních dat, spolupráce se zahraničím začala již po roce 2000 a stále pokračuje.

2 DEFINICE CÍLŮ A PRINCIPŮ NÁRODNÍ STRATEGIE

Základním cílem pro příštích pět let by mělo být vytvoření podmínek pro snížení bariér, podporu spolupráce a diverzifikaci přístupu k dlouhodobé ochraně a zároveň by měly být jasně stanoveny standardy pro tuto oblast.

K tomuto cíli bychom se měli dostat postupným rozvíjením aktivit v jednotlivých oblastech:

- **Rozvoj plánování dlouhodobé ochrany** – na úrovni institucí písemně stanovené plány dlouhodobé ochrany digitálních dat, které budou obsahovat měřitelné ukazatele; strategické plánování na úrovni systému knihoven a jednotlivých projektů.
- **Zlepšení organizačních podmínek** – pro dlouhodobou ochranu v knihovnách a to jak na úrovni jednotlivých knihoven (standardizace procesů a správy dat, plánování a dokumentace vyžaduje další oddělení, pracovní síly, finance), tak na úrovni systému knihoven (například vznikem koordinačního centra, viz níže).
- **Podpora spolupráce a sdílení –** jak na domácí úrovni mezi paměťovými institucemi různých resortů, s vysokými školami a výzkumnými organizacemi atd., tak na mezinárodní úrovni s jinými paměťovými institucemi. Sdílení procesů, plánů, metodik, hardwarové i softwarové infrastruktury[21] včetně návrhu řešení kontinuity dat po zániku instituce.
- **Snížení bariér** – zajistit, aby každá instituce, která vytváří nebo shromažďuje digitální data, měla přístup k znalostem, službám nebo konkrétním technologiím. Manažer každé instituce musí mít takové znalosti nebo dokumentaci, aby mohl rozhodnout, jak se bude dlouhodobou ochranou zabývat, jak budou data ukládána, a měl možnost využít služeb systémů jiných institucí nebo státem poskytovaných služeb.
- **Rozvoj mechanismů financování dlouhodobé ochrany** – vytvoření udržitelného a kontrolovatelného finančního nástroje pro financování projektů v oblasti dlouhodobé ochrany, a to včetně jednorázových akcí. Mechanismus by měl rozlišit prostý nákup hardwaru, zavádění systémů pro správu dat, jednorázové akce ochrany (přesun z jednoho hardwaru na jiný, převody mezi formáty), komplexní logickou dlouhodobou ochranu a výzkumné projekty.
- **Používání (zahraničních) modelů pro stanovení nákladů na dlouhodobou ochranu** (*digital preservation costs*) – vytvoření obecného doporučené měření nákladů na dlouhodobé ukládání digitálních dat s ohledem v kontextu financování Českých knihoven,
- **Podpora vzdělávání, kvalifikace a výzkum** – existující formální a jakým způsobem má být ochrana zajištěna, tj. jaké jsou požadavky na systémy, ve kterých mají tato data být ukládána vysokoškolské obory typu Knihovnictví a informační věda, Archivnictví a Muzeologie a některé v oblasti informačních technologií by do svých učebních programů měly začlenit problematiku dlouhodobé ochrany digitálních dat, případně vytvořit nová zaměření specializovaná na správu a dlouhodobou ochranu digitálních dat

[21] Jedná se o projekty typu CLOCKSS: *The CLOCKSS Archive: A Trusted Community-Governed Archive* [online]. South Orange (NJ, USA): CLOCKSS, c2015 [cit. 2016-12-05]. Dostupné z: https://www.clockss.org/clockss/Home

např. podle doporučení projektu DigCurV[22] nebo podle již existujících učebních programů v zahraničí. Je také žádoucí specifická podpora projektů spolupráce a výzkumu v oblasti dlouhodobé ochrany (knihovny a vysoké školy), systematická podpora studentských stáží a vysokoškolských kvalifikačních prací.

- **Metodické centrum pro dlouhodobou ochranu digitálních dat** – mělo by plnit role v oblasti zajištění kvality dlouhodobé archivace. Cílem by mělo být zavedení národního mechanismu externí certifikace důvěryhodného dlouhodobého úložiště pro data v knihovnách. To musí být podpořeno fungujícím systémem a metodikou tzv. sebehodnocení, např. dle DSA (Data Seal of Approval[23]) nebo Nestor Seal[24]. Centrum by mělo vytvořit doporučení,, jak mají instituce postupovat, chtějí-li dosáhnout statutu důvěryhodného dlouhodobého úložiště.
- **Důraz na prosazování dlouhodobé ochrany** – pro úspěch dlouhodobé ochrany jsou klíčové jak finanční prostředky, tak odhodlání lidí se touto problematikou zabývat. Protože dlouhodobá ochrana je finančně náročná činnost probíhající v pozadí, často daleko od každodenního provozu digitální knihovny, je třeba tyto činnosti popularizovat a vysvětlovat jejich význam pro zachování služeb knihoven.

[22] *DigCurV: Digital Curator Vocational Education Europe* [online]. DigCurV, c2010 [cit. 2016-12-05]. Dostupné z: http://www.digcur-education.org/

[23] *Data Seal of Approval* [online]. [cit. 2016-12-05]. Dostupné z: http://www.datasealofapproval.org/en/

NESTOR CHECKLISTS: Selection of Solutions and Components for Digital Long-Term-Preservation nestor-Checklist by AG Kooperation & Vernetzung [online]. Frankfurt am Main: Deutsche Nationalbibliothek, Last update: 30.03.2012 [cit. 2016-12-05]. Dostupné z: http://www.dnb.de/Subsites/nestor/EN/Publikationen/Checklisten/checklisten_node.html

3 IDEÁLNÍ STAV ZA PĚT LET V OBLASTI DLOUHODOBÉ OCHRANY

Existuje systém pro dlouhodobou ochranu knihovních dat. Nejedná se o systém v technickém slova smyslu, ale o soubor pravidel, institucí, technických zařízení a aplikací, vzdělávacích programů a jejich organizační rámec. Systém může být součástí širšího pojetí ochrany digitálních dat v ČR, nejen dat z knihoven, ale také např. z archivů, muzeí, galerií, vědeckých institucí apod. Jeho hlavními složkami jsou hardwarová a softwarová řešení, instituce, finanční mechanismy a standardy.

- Hardwarová a softwarová řešení zahrnují:
  - garantované datové centrum[25] pro zajištění bitové ochrany digitálních dat knihoven;
  - garantovaný LTP systém[26] pro zajištění logické ochrany digitálních dat knihoven;
  - lokální řešení (otevřené nebo komerční) jednotlivých institucí různé úrovně a kvality;
  - datové úložiště CESNET (případně výpočetní centra vysokých škol).
- Institucemi jsou:
  - Metodické centrum pro dlouhodobou ochranu digitálních dat;
  - centrální datové centrum jako samostatný subjekt;
  - jednotlivé knihovny (případně i vysoké školy, archivy či muzea).
- Finanční mechanismy zahrnují:
  - stávající mechanismy (např. program VISK 7) obohacené o nové prvky;
  - nové finanční mechanismy financující aktivity v oblasti dlouhodobé ochrany, vlastní aktivity institucí nebo využití centrální infrastruktury;
  - využití modelů pro stanovení nákladů na dlouhodobou ochranu jako nástroje pro plánování provozu a financování úložišť.
- Standardy jsou míněny:
  - standardy pro data a metadata;
  - doporučení a metodiky pro dlouhodobou ochranu;
  - mezinárodní normy.

3.1 Jádro systému pro dlouhodobou ochranu knihovních dat

Jádro systému pro dlouhodobou ochranu knihovních dat tvoří garantovaná datová centra[27].

[25] Datové centrum by optimálně mělo být garantované státem.
[26] LTP systém by optimálně měl být garantován státem.
[27] Datová centra by optimálně měla být garantována státem a poskytována např. MK či sdružením CESNET.

- Datová centra mají více technologií ukládání k zajištění řízení.
- Infrastruktura umožňuje distribuci dat mezi oblastmi; data jsou dostupná prostřednictvím rychlého datového připojení (nejedná se o převážení pásek); infrastruktura umožňuje flexibilní nabídku služeb a instituce nakupují nebo využívají kapacity podle potřeb.
- Existuje dohoda o automatickém převzetí odpovědnosti za data vznikající za veřejné finance - pokud některá instituce vytváří digitální data určená k dlouhodobé ochraně a přitom dlouhodobou ochranu dat neřeší vlastními kapacitami, musí mít možnost jejich ochranu řešit jiným způsobem. Musí mít možnost data k dlouhodobé ochraně předat a musí vědět komu, jak a v jaké podobě. Podobně musí být jasně stanoveno, kdo bude za uložená data odpovědný v případě zániku instituce.
- Pro data kterékoli knihovny je zajištěna minimálně ochrana na bitové úrovni – každá instituce má pro zajištění bezpečeného uložení dat možnost využít kapacitu a služby CESNET nebo garantovaného datového centra[28]. Národní cloud by mohl mít podobu IaaS (*Infrastructure as a Service*), resp. SaaS (*Storage as a Service*) atd.

3.2 Další certifikované systémy

Některé z velkých institucí nebo sdružení (NK ČR, MZK, KNAV, kraje, vysoké školy, např. MU či UK, CESNET) mají k dispozici certifikovaná řešení pro dlouhodobou ochranu digitálních dat včetně odpovídajícího financování a personálu.

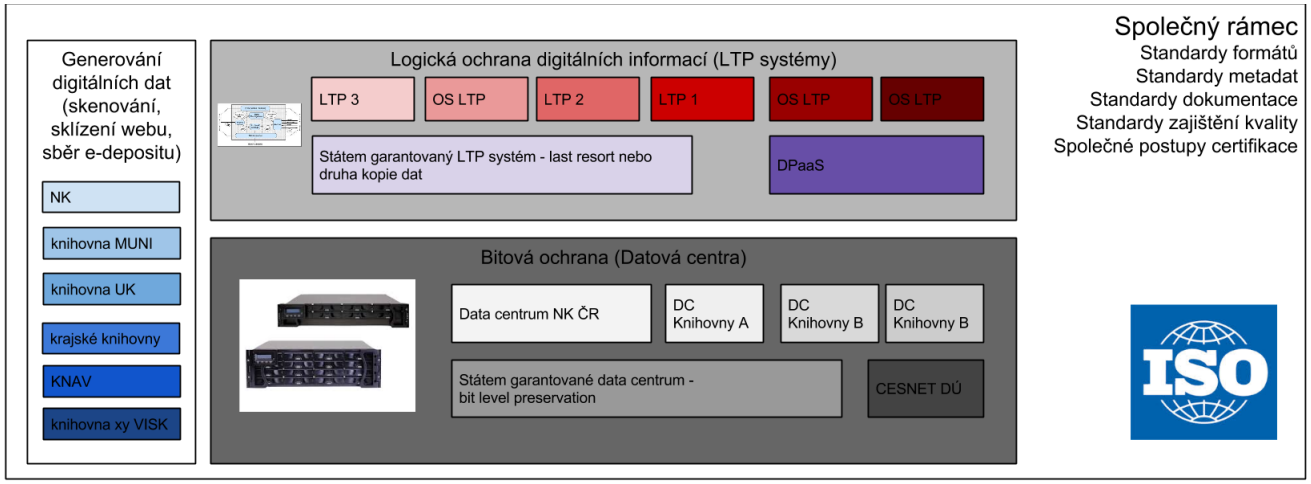
- Existující LTP systémy mají povinnost převzít dat od menších institucí v případě jejich zániku a mají smlouvy mezi sebou o tomtéž.
- Standardizace formátů a metadatových specifikací napříč institucemi stejného typu (i různé velikosti).
- Standardizace pracovních postupů a dokumentace.
- Některé instituce používají externí cloudové úložiště nebo služby cloudového LTP systému[29].
- Některé z institucí poskytují přímo služby dlouhodobé ochrany jiným institucím, tj. DPaaS (*Digital Preservation as a Service*).

3.3 Otevřená řešení

Vedle vždy existující možnosti pořízení komerčního LTP systému existuje státem podporované a snadno dostupné otevřené řešení[30] systému na dlouhodobou ochranu pro malé a střední instituce. Instituce mají

[28] Datové centrum by optimálně mělo být garantovanáno státem a poskytováno MK či sdružením CESNET.
[29] Tím je např. LTP systém Preservica ve své cloudové verzi: PRESERVICA. *Preservica Editions* [online]. Boston (MA, USA): Preservica, c2016 [cit. 2016-12-05]. Dostupné z: http://preservica.com/editions-pricing/
[30] Implementace otevřeného řešení neznamená nutně, že každá malá instituce by měla provozovat LTP systém v plném rozsahu všech funkcí OAIS. Klíčové funkce v oblasti plánování ochrany, sledování technologií, vytváření

na výběr – buď využijí existující otevřené řešení, nebo mohou využít službu některého z provozovaných systémů ve větších institucích. Minimálně ochrana na úrovni bitů zaručena pro všechna data vznikající v projektech financovaných z veřejných zdrojů[31].



Obr. 1: Složky národního systému pro ochranu digitálních dat

3.4 Data dostupná výhradně v digitální podobě (born-digital data)

- NK ČR a další zákonem pověřené knihovny mají ověřený mechanismus na získávání a zpracování born-digital dokumentů na pevných nosičích cestou povinného výtisku a zajišťují jejich logickou dlouhodobou ochranu.
- NK ČR a další pověřené knihovny pravidelně sbírají obsah českého internetu (doména .cz a další weby bohemikální povahy) a zajišťují logickou dlouhodobou ochranu těchto dat. Webarchiv prochází pravidelně auditem kvality podle příslušné ISO normy.

3.5 Vzdělávání, výzkum a certifikace

- Součástí infrastruktury je Metodické centrum pro dlouhodobou ochranu digitálních dat, které poskytuje vzdělávání a poradenství a připravuje standardy a metodiky.
- Vysokoškolské vzdělávací programy jsou odpovídajícím způsobem obohaceny o problematiku dlouhodobé ochrany digitálních dat.
- Dlouhodobá ochrana digitálních dat je stálou, reálně prováděnou a věcně, metodicky i finančně podporovanou agendou MK a dalších ministerstev, v souladu s doporučeními EU, a to nejen na papíře.

---

standardů atd. by měla nadále vykonávat NK ČR ve spolupráci s ostatními knihovnami. To ale nevylučuje, že některé menší instituce mohou usilovat o certifikaci DSA apod.

[31] Jedná se úroveň poskytovatele dotace či státu.

- V definicích pracovních pozic ve veřejných institucích jsou jasně popsány pracovní pozice a kvalifikační požadavky na kurátory digitálních dat, správce digitálních knihoven a metadatové specialisty.

3.6 Sdílení a financování

- MK ČR nabízí dlouhodobě udržitelné mechanismy financování dostupné pro projektové aktivity, vzdělávání v oblasti dlouhodobé ochrany a LTP systémů; experimenty a hodnocení vznikajících LTP systémů, nástrojů atd.
- Programová spolupráce knihoven je koordinována v rámci sektoru a s dalšími paměťovými institucemi, jako jsou muzea, galerie apod.

4 KONKRÉTNÍ DOPORUČENÁ OPATŘENÍ

4.1 Plán na zajištění dlouhodobé ochrany digitálního kulturního a vědeckého dědictví v ČR

- Stanovit cíle, organizační opatření, pravidla, přidělit kompetence (bitová ochrana, logická ochrana digitálních dat) pro oblast knihoven.
- Zpracovat strategii MK, která by měla:
    - stanovit, co jsou digitální kulturní a vědecká data trvalé hodnoty.
    - stanovit, co je dlouhodobá ochrana digitálních dat a jaké procesy budou využity k jejímu zajištění.
    - odvolat se na české verze norem ISO 16363 a ISO 14721 a na normu ISO/TR 14873, zmíněny by měly být i normy ISO/IEC 27000 a ISO 9000 a další standardy (DSA, Nestor Checklist).
- Nastavit program financování aktivit spojených s trvalou ochranou digitálních kulturních a vědeckých dat tak, aby bylo možné využít existující infrastruktury (CESNET, NK ČR) a zachovat možnost institucí rozhodovat o tom, jak budou jejich data ochraňována. Závazkem pro získání financí by měl být postup k certifikaci, resp. snaha vyhovět požadavkům DSA i zajistit kvalitu procesů v organizaci (ISO 9001). To mj. vyžaduje specifikaci procesu a provádění certifikace digitálních úložišť v institucích v gesci MK.

4.2 Garantovaná datová centra pro zajištění bitové ochrany digitálních dat

- Vytvořit podmínky pro to, aby všechny instituce disponující velkými objemy dat mohly získat pokročilé řešení pro zajištění bitové ochrany digitálních dat formou služby – tj. garantované datové centrum[32] (nebo více datových center).[33] Datové centrum musí splňovat standardy ISO/IEC 27000 a další standardy. Toto datové centrum by mělo sloužit jako sekundární lokalita pro kteroukoliv instituci nebo primární úložiště pro menší instituce. Musí mít zajištěno trvalé financování.

4.3 Garantovaný systém pro zajištění logické ochrany digitálních dat

- CESNET nebo nově budované datové centrum MK[34] by mohly provozovat tzv. LTP systém a s jeho pomocí začít nabízet certifikované řešení pro logickou dlouhodobou ochranu digitálních dat jako

[32] Datové centrum by měl optimálně garantovat stát.
[33] Služby datového centra by mohl poskytovat např. CESNET.
[34] Systém by měl optimálně v obou variantách garantovat stát.

službu (DPaaS) pro instituce v gesci MK.

- Datové centrum nebo datová centra by prostřednictvím LTP systému měla poskytovat služby dlouhodobé ochrany vyšší úrovně, tj. úrovně logické. Financován by měl být přenos dat do těchto datových center, s tím, že by paralelně neměla klesnout podpora lokálně provozovaných řešení (redundance, technologická odlišnost).
- LTP systém poskytovaný jako služba by měl splňovat standardy ISO 16363 a ISO 14721 prokázané zveřejněným interním auditem a externí mezinárodní certifikací. Služby by měly používat především instituce, které nechtějí samy LTP systémy budovat, obvykle tedy menší instituce.
- Aby takové řešení mohlo existovat jako dlouhodobé důvěryhodné úložiště[35], potřebuje především garantované dlouhodobé financování. Bez zajištěného financování nelze očekávat, že řešení získá externí certifikaci podle ISO 16363.

4.4 Postupné budování sítě certifikovaných úložišť v jednotlivých institucích

- MK by ve svých finančních mechanismech mělo vyčlenit finance na pořízení, zavedení a certifikaci LTP systémů (např. podle DSA). To bude vyžadovat také další standardizaci procesů v jednotlivých institucích, převody a čištění dat a bude to znamenat také další náklady.
- Financování musí být zajištěno s oporou ve státním rozpočtu a rozpočtových kapitolách nejlépe na delší dobu dopředu.
- Jedním z řešení financovaných z tohoto zdroje by mělo být rozvíjení řešení NDK na úroveň plnohodnotného LTP systému[36].
- Instituce by měly by mít možnost si vybrat mezi službou zmíněného datového centra vybudovaného pro účely bitové ochrany v části 4.2 a svým lokálním řešením. V datovém centru by v takovém případě mělo být možné i např. uložit pouze zálohovou kopii dat.

4.5 Podpora otevřených řešení pro zajištění logické ochrany digitálních dat

- Pro potřeby českých kulturních institucí by měl by být vytvořen dostupný otevřený LTP systém.
- Malé nebo střední instituce by mohly systém používat pro správu a archivaci svých digitálních fondů.
- Mělo by se jednat o řešení založené na ověřených otevřených komponentech, které jsou úspěšně

[35] Dlouhodobé důvěryhodné úložiště je v tomto dokumentu též označováno jako LTP systém.
[36] NK ČR rozvíjí a aktualizuje metadatový profil (vnitřní metadatový formát) pro archivní balíčky v LTP úložišti NDK tak, aby byl tvořen na základě uznávaných mezinárodních metadatových formátů (např. PREMIS) a reflektoval jejich změny.

nasazovány v zahraničí, např. Fedora Repository[37], Archivematica[38], případně RODA[39] – výhodou je mezinárodní komunita, neustálý rozvoj, komunitní podpora apod.

Takové řešení by mělo v příštích letech vzniknout jako výstup z probíhajícího projektu ArcLIB, který je realizován v rámci programu NAKI II, a to KNAV, MZK, MU a NK ČR.

4.6 Zřízení Metodického centra pro dlouhodobou ochranu digitálních dat

Zřídit a formalizovat fungování Metodického centra pro dlouhodobou ochranu digitálních dat např. v návaznosti na plánované Metodické centrum digitálních knihovních dokumentů jako součást Plánu implementace Státní kulturní politiky na léta 2015 –2020[40].

Metodické centrum pro dlouhodobou ochranu digitálních dat by mělo:

- být garantováno MK;
- vytvářet metodiky, národní strategie, národní standardy;
- vést vývoj „národních systémů nebo řešení";
- podporovat kulturní instituce provádějící nebo začínající s dlouhodobou ochranou digitálních dat;
- spolupracovat se zahraničními odborníky na oblast dlouhodobé archivace a jako svébytná entita se účastnit např. evropských projektů;
- zaštiťovat a podporovat proces certifikace digitálních úložišť, překlady relevantních norem a metodik;
- být centrem plánování financování činností vztahujících se k dlouhodobé archivaci na národní úrovni – vypisování a správa grantů aj.;
- zveřejňovat výsledky své činnosti, informovat o své činnosti vládu ČR apod.;
- spolupracovat s vysokými školami na přípravě a realizaci učebních programů;
- ve spolupráci s provozovatelem centrálního datového centra propagovat jeho služby.

4.7 Krajské knihovny jako producenti a správci digitálních dat

- Krajské knihovny by měly mít jasné plány a možnosti na uložení dat a jejich logickou ochranu.
- K dispozici jim bude metodika, podpora, strategie z Metodického centra pro dlouhodobou

---

[37] http://fedorarepository.org/
38 *Archivematica: Main Page* [online]. Archivematica, This page was last modified on 18 September 2016, at 12:10 [cit. 2016-12-05]. Dostupné z: https://www.archivematica.org/wiki/Main_Page
39 *RODA Community: Repository of Authentic Digital Objects* [online]. RODA Community, c2012 [cit. 2016-12-05]. Dostupné z: http://www.roda-community.org/
[40] *Plán implementace Státní kulturní politiky na léta 2015 – 2020* [online]. [Praha: MK ČR, 2015] [cit. 2016-12-05]. Dostupné z: https://www.mkcr.cz/doc/cms_library/plan-implementace-3862.docx

ochranu digitálních dat.

- Měly by mít možnost získat financování pro procesy dlouhodobé ochrany.
- Předpokládá se, že pro bitovou ochranu využijí krajské knihovny úložnou kapacitu krajských technologických center provozovaných v rámci jednotlivých krajských úřadů. Měly by rovněž mít možnost využít služeb garantovaných datových center[41].

4.8 Malé instituce jako producenti dat

- Malé instituce, např. digitalizující v malém rozsahu vlastními silami, mají mít možnost využití datových center (viz bod 4.2 koncepce) či využití otevřeného LTP systému (provozovaného krajskou knihovnou či vlastními silami).
- K dispozici jim bude metodika, podpora, strategie z Metodického centra pro dlouhodobou ochranu digitálních dat.
- Měly by mít možnost získání financování pro procesy dlouhodobé ochrany[42], např. z programu VISK.
- Program VISK by měl zahrnovat popis standardů dlouhodobé ochrany dat – doporučení vycházející z normy ISO 14721 a dalších metodik, a to s jasně popsanými možnostmi řešení zajištění logické ochrany dat.

4.9 Finanční mechanismy

- Optimálně financování ochrany digitálního kulturního dědictví v rámci některého z programů TAČR pro relevantní ministerstva (MK, MV, MŠMT).
- Instituce musejí mít možnost požádat o několikaleté financování s následnou možností pokračování podpory na projekty uložení dat a jejich logické ochrany (ne vytváření dat). Nemělo by se jednat o financování nových pracovních pozic, prostředky musejí být získány reorganizací a evaluací stávajících procesů a aktivit v organizaci. Může se např. jednat o projekty na:
    - instalaci LTP systému,
    - pořízení hardwarové infrastruktury (v odůvodněných případech, kdy není možné HW sdílet nebo využít IaaS nebo DPaaS),
    - převody dat do nových systémů, na nová média, ochranné akce na velkém množství dat aj.
- Program VISK by měl obsahovat nový podprogram týkající se dlouhodobé archivace. Měl by podpořit zavedení otevřeného nebo komerčního LTP systému, zavedení služby LTP a/nebo

[41] Datová centra by měla být optimálně garantována státem.
[42] Např. na konsolidaci svých dat, správu, validaci a konverzi dat či na vytvoření archivních balíčků.

spojené náklady (příprava dat, převod dat, kontroly dat, realizace přenosu dat atd.).

- Financování digitalizace z programu VISK by mělo být podmíněno prokázáním schopnosti zajistit dlouhodobou archivaci vytvořených dat. Každý žadatel by měl předložit plán archivace, v němž přesně stanoví, jak budou data dlouhodobě uchovávána. Bez schopnosti zajistit dostatečně bitovou ochranu výsledků digitalizace by nemělo být financování přiděleno[43]. Program VISK by měl tuto ochranu financovat, stejně jako financuje pořízení dat.
- Knihovny využívají modely pro stanovení nákladů na dlouhodobou ochranu jako nástroje pro plánování provozu a financování svých úložišť.

4.10 Organizační a systémové změny

- Zřizovací listiny jednotlivých organizací musí výslovně obsahovat závazek logické dlouhodobé ochrany digitálních dat.
- Stávající zákony pokrývající tzv. povinný výtisk musejí být upraveny tak, aby bez jakýchkoliv pochyb zahrnovaly povinnost odevzdávat i dokumenty dostupné výhradně v digitální podobě (born-digital konkrétním knihovnám prostřednictvím konkrétních mechanismů.
- Národní soustava povolání by měla být aktualizována v souladu s rozvojem oblasti správy digitálních dat v knihovnách a dalších paměťových institucích. Zejména ve spolupráci s odborníky z archivů, případně dalších odvětví, by měl být zpracován společný kvalifikační standard pro povolání kurátor digitálních dat. Při tvorbě nových specializací a aktualizaci existujících kvalifikačních standardů je především třeba zohlednit vývoj specializace správce digitální knihovny a katalogizátor směrem ke kompetencím potřebným pro správu LTP systému digitálních knihoven (např. metadatové standardy, informatické kompetence).

4.11. Konsolidace dříve pořízených dat a interoperabilita

- Řada institucí pracuje s daty, která nejsou opatřena dostatečnými technickými metadaty a neodpovídají požadavkům na informační balíčky pro dlouhodobé ukládání[44]. Jedná se o historická data, data nad rámec současného standardu NDK nebo data, která existují pouze v systémech digitálních knihoven apod. I tato data je třeba konsolidovat, opatřit dostatečnými metadaty a připravit na archivaci v systémech, které se budou snažit o zajištění maximální kompatibility se standardy NDK.

[43] Instrukce by měla stanovit minimální požadavky na ochranu dat – umístění ve dvou lokalitách na dvou typech médií dostupných online, jedna offline záloha navíc.
[44] Např. SIP balíčky.

5 DOPORUČENÍ V OBLASTI ZAJIŠTĚNÍ KVALITY, STANDARDIZACE A CERTIFIKACE

5.1 Zajištění kvality; audit a certifikace

- Každá instituce, která sama provádí dlouhodobou ochranu nebo pro kterou tuto činnost vykonává externí subjekt, musí pravidelně procházet auditem. Úspěšný audit by měl být podmínkou pro získání (nebo pokračování) financování. Výstupem z auditu může být certifikace podle některé z níže uvedených norem.
- Procesy v organizacích, které se zabývají dlouhodobou ochranou, musejí být zdokumentované a musejí odpovídat standardům. Ideální by bylo, kdyby Metodické centrum pro dlouhodobou ochranu digitálních dat Metodické centrum pro dlouhodobou ochranu digitálních dat formulovalo požadavky na zajištění kvality a zajistilo externí audit. Dodržování procesů, standardů, metodik by také mělo být podmínkou pro získání financování z veřejných zdrojů. Dokumentace a popisy procesů musejí být veřejně přístupné a otevřené, stejně jako např. technické řešení nebo smlouvy o outsourcování služeb.
- Základním orientačním bodem pro metodiku interního i externího auditu musejí být normy ISO 16363 a ISO 14721 s tím, že realizace tzv. sebehodnocení nebo sebehodnocení pod dohledem odborníků na danou oblast (peer reviewied self-auditu) by měla probíhat podle národní metodiky stanovené Metodickým centrem pro dlouhodobou ochranu digitálních dat. V organizacích je také třeba také pracovat na splnění požadavků dalších souvisejících norem (ISO/IEC 27001, ISO 9001).
- Autoritou provádějící audity by mělo být Metodické centrum pro dlouhodobou ochranu digitálních dat, které by disponovalo organizační platformou a možností najímat kvalifikované auditory.

Získání oficiální certifikace podle ISO 16363 od akreditované certifikační autority není u všech

- projektů nezbytné ani potřebné. Výjimku by mohly tvořit velké projekty financované EU, případně národních projekty. Oficiální externí certifikace od certifikační autority bude poměrně finančně[45] i časově náročná. Pravidelné sebehodnocení nebo externí audit podle pravidel dohodnutých na národní úrovni, spolu s publikací veškeré dokumentace k sebehodnocení a hodnocení auditorů, bude pro většinu projektů dostačující zárukou snahy o dosažení statutu důvěryhodného dlouhodobého úložiště ve smyslu normy ISO 16363.
- Při zavádění a provozování LTP systému je třeba postupovat podle společného evropského rámce

[45] Proces certifikace dle ISO 16363 podle prováděcí normy ISO 16919 bude vyžadovat auditorskou firmu, která bude akreditována u certifikační autority, tj. u Českého institutu pro akreditaci, a.s. (http://www.cia.cz/). Tato firma bude podle ISO 16919 mít nejen akreditaci k ISO 16363, ale také k ISO 17021.

pro certifikaci[46], který doporučuje postup ve třech krocích – sebehodnocení podle DSA, sebehodnocení podle ISO 16363 a externí audit podle ISO 16363. Nástroje jako DRAMBORA[47] nebo PLATTER[48] by měly sloužit při návrhu systému a analýze rizik, nebo pro metodickou orientaci v oblasti LTP. Je možné použít i německé ekvivalenty ISO 16363 a DSA (Nestor seal a Nestor criteria catalogue – DIN 31644:2012), nicméně národní metodické centrum pro dlouhodobou ochranu digitálních dat nemusí garantovat stejnou míru podpory (standardy pro dokumentaci a doporučení, asistenci při sebehodnocení atd.) pro obě skupiny norem.

5.2 Doporučené postupy

Metodické centrum pro dlouhodobou ochranu digitálních dat by mělo také vypracovat doporučení pro:

- distribuci kopií dat mezi technologie a lokality,
- dokumentaci (řízení změn, řízení procesů, správy nastavení),
- zajištění kvality řízení a organizace práce,
- standardy vztahující se k softwaru a worfklow pro zpracování dat podle ISO 14721 .

[46] *TrustedDigitalRepository.eu: Making the digital world more reliable* [online]. [cit. 2016-12-05]. Dostupné: http://www.trusteddigitalrepository.eu/Trusted%20Digital%20Repository.html
[47] *DRAMBORA interactive* [online]. Digital Curation Centre, Date Published: 1 February 2008, Last Modified: 6 August 2015 [cit. 2016-12-05]. Dostupné: http://www.repositoryaudit.eu/
[48] ROSENTHAL, Colin et al. *Průvodce plánem důvěryhodného digitálního repozitáře (PLATTER)*. Praha: Národní knihovna ČR, 2009. ISBN 978-80-7050-569-4. Dostupné také z: http://www.ndk.cz/platter-cz

6 POSTUP REALIZACE KONCEPCE

Po schválení koncepce by se její realizací měla zabývat pracovní skupina Ústřední knihovnické rady MK, která by navázala na výstupy práce pracovní skupiny v rámci plnění Priority č. 2 Koncepce rozvoje knihoven ČR na léta 2011 – 2015[49].

6.1 Optimální postup realizace koncepce

První fáze (2016)

První fází je analýza a pilotní využití otevřených systémů.

Druhá fáze (2017–2018)

Druhá fáze je tvořena následujícími kroky:

- vytvoření Metodického centra pro dlouhodobou ochranu digitálních dat;
- zřízení garantovaného datového centra pro zajištění bitové ochrany digitálních dat knihoven;
- vytvoření metodik pro LTP (bod 5.1 a 5.2);
- vybudování systému pro výběr elektronického povinného výtisku;
- úpravy finančních mechanismů (mj. programu VISK) tak, aby zahrnovaly problematiku archivace digitálních dat (bez plánu archivace nelze digitalizovat);
- vytvoření nových finančních mechanismů pro pořízení a budování LTP systémů;
- vytvoření doporučení pro zajištění kvality v oblasti dlouhodobé archivace (normy, standardy, certifikace);
- asistence Metodického centra pro dlouhodobou ochranu digitálních dat při certifikaci;
- prosazení standardů kvality archivace webu;
- tlak na instituce, aby se věnovaly zajištění kvality podle ISO 9000 a ISO/IEC 27000;
- úprava knihovního zákona a zákonů týkajících se povinného výtisku a navazujících prováděcích předpisů;
- obohacení Národní soustavy povolání o relevantní pozice.

Třetí fáze (2019–2020)

Třetí fáze je tvořena následujícími kroky:

- vybudování garantovaného systému pro logickou ochranu dat, který bude schopen nabízet logickou ochranu dat jako službu;

---

[49] *Plnění priorit a úkolů Koncepce rozvoje knihoven ČR na léta 2011 – 2015* [online]. [Ústřední knihovnická rada] [cit. 2016-12-05]. Dostupné z: http://files.u-k-r.webnode.cz/200000220-a1c3da3a54/plneni_koncepce_2011_2015.docx

- financování budování lokálních řešení nebo využití služeb pro logickou ochranu digitálních dat;
- financování podpory využití wotevřených systémů pro logickou ochranu;
- financování certifikace činností vztahujících se k digitální archivaci.

6.2 Minimální očekávaný stav v roce 2020

V roce 2020 by měly být dokončeny následující činnosti:

- analýza a pilotní využití otevřených systémů;
- vytvoření Metodického centra pro dlouhodobou ochranu digitálních dat;
- změny knihovního zákona a prováděcí předpis;
- vytvoření metodik pro digitální archivaci (bod 5.1 a 5.2);
- vybudování systému pro výběr elektronického povinného výtisku;
- úpravy finančních mechanismů (mj. programu VISK) tak, aby zahrnovaly problematiku archivace digitálních dat (bez plánu archivace nelze digitalizovat);
- vytvoření doporučení pro zajištění kvality v oblasti dlouhodobé archivace (normy, standardy, certifikace);
- asistence Metodického centra pro dlouhodobou ochranu digitálních dat při certifikaci;
- úprava knihovního zákona a zákonů týkajících se povinného výtisku;
- obohacení Národní soustavy povolání o relevantní pozice;
- financování certifikace, auditu či sebehodnocení činností vztahujících se k dlouhodobé archivaci.

7 SEZNAM POUŽITÝCH ZKRATEK

Zkratka	Rozpis
ALTO	Analyzed Layout and Text Object
CDA	Centrálny dátový archiv
DANS	Data Archiving and Networking Services
DPaaS	Digital Preservation as a Service
DSA	Data Seal of Approval
DTD	Document Type Definition
EU	Evropská unie
FAMU	Filmová a televizní fakulta Akademie múzických umění v Praze
GESIS	Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen
IaaS	Infrastructure as a Service
JPEG	Joint Photographic Experts Group
KNAV	Knihovna AV ČR, v. v. i.
LTP	Long-term Preservation
METS	Metadata Encoding & Transmission Standard
MK	Ministerstvo kultury
MŠMT	Ministerstvo školství, mládeže a tělovýchovy
MU	Masarykova univerzita
MV	Ministerstvo vnitra
NA	Národní archiv
NDK	Národní digitální knihovna
NFA	Národní filmový archiv
NK ČR	Národní knihovna České republiky
NTK	Národní technická knihovna
OAIS	Open Archival Information System
OPIS	Operačný program Informatizácia spoločnosti
PREMIS	PREservation Metadata: Implementation Strategies
SaaS	Storage as a Service

TAČR	Technologická agentura ČR
TIFF	Tagged Image File Format
URN:NBN	Uniform Resource Name: National Bibliography Number
VISK	Veřejné informační služby knihoven
VŠCHT	Vysoká škola chemicko-technlogická v Praze

8 SEZNAM CITOVANÝCH NOREM

Pozn.: Na první úrovni jsou uvedeny záznamy norem (včetně stručných údajů o předchozích vydáních norem, popř. přípravě vydání nových), na druhé úrovni jsou pak u mezinárodních norem zachyceny údaje o jejich překladech do češtiny.

- DIN 31644:2012. *Information und Dokumentation – Kriterien für vertrauenswürdige digitale Langzeitarchive*. Berlin: DIN, 2012.
- ISO 14721:2012. *Space data and information transfer systems – Open archival information system (OAIS) – Reference model*. Geneva: ISO, 2012. Nahrazuje a ruší ISO 14721:2003. Dostupné také z: https://public.ccsds.org/pubs/650x0m2.pdf (původní publikace CCSDS 650.0-M-2).
    - ČSN ISO 14721:2014. *Systémy pro přenos dat a informací z kosmického prostoru – Otevřený archivační informační systém – Referenční model*. Praha: ÚNMZ, 2014.
- ISO/TR 14873:2013. *Information and documentation – Statistics and quality issues for web archiving*. Geneva: ISO, 2013.
    - české vydání není k dispozici
- ISO 16363:2012. *Space data and information transfer systems – Audit and certification of trustworthy digital repositories*. Geneva: ISO, 2012. Dostupné také z: https://public.ccsds.org/pubs/652x0m1.pdf (původní publikace CCSDS CCSDS 652.0-M-1).
    - ČSN ISO 16363:2014. *Systémy pro přenos dat a informací z kosmického prostoru – Audit a certifikace důvěryhodných digitálních úložišť*. Praha: ÚNMZ, 2014.
- ISO 16919:2014. *Space data and information transfer systems – Requirements for bodies providing audit and certification of candidate trustworthy digital repositories*. Geneva: ISO, 2014.
    - české vydání není k dispozici
- ISO 17021-1:2015. *Conformity assessment – Requirements for bodies providing audit and certification of management systems – Part 1: Requirements*. Geneva: ISO, 2015. Nahrazuje a ruší ISO/IEC 17021:2011.
    - ČSN EN ISO/IEC 17021-1:2016. *Posuzování shody – Požadavky na orgány poskytující služby auditů a certifikace systémů managementu – Část 1: Požadavky*. Praha: ÚNMZ, 2016.
- ISO/IEC 27000:2016. *Information technology – Security techniques – Information security management systems – Overview and vocabulary.* Geneva: ISO/IEC, 2016. Nahrazuje a ruší ISO/IEC 27000:2014. Dostupné také z: http://standards.iso.org/ittf/PubliclyAvailableStandards/c066435_ISO_IEC_27000_2016(E).zip (v angličtině) a http://standards.iso.org/ittf/PubliclyAvailableStandards/c066435_ISO_IEC_27000_2016(F).zip (ve francouzštině).
    - ČSN ISO/IEC 27000:2014. *Informační technologie – Bezpečnostní techniky – Systémy řízení bezpečnosti informací – Přehled a slovník*. Praha: ÚNMZ, 2014.

- ISO/IEC 27001:2013. *Information technology – Security techniques – Information security management systems – Requirements*. Geneva: ISO/IEC, 2013. Nahrazuje a ruší ISO/IEC 27001:2005.
    - ČSN ISO/IEC 27001:2014. Informační technologie – Bezpečnostní techniky – Systémy řízení bezpečnosti informací – Požadavky. Praha: ÚNMZ, 2014.
- ISO 28500:2009. *Information and documentation – WARC file format*. Geneva: ISO, 2009. Nové vydání normy (ISO/DIS 28500) je v přípravě.
    - české vydání není k dispozici
- ISO 9000:2015. *Quality management systems – Fundamentals and vocabulary*. Geneva: ISO, 2015. Nahrazuje a ruší ISO 9000:2005.
    - ČSN EN ISO 9000:2016. *Systémy managementu kvality – Základní principy a slovník*. Praha: ÚNMZ, 2016.
- ISO 9001:2015. *Quality management systems – Requirements*. Geneva: ISO, 2015. Nahrazuje a ruší ISO 9001:2008.
    - ČSN EN ISO 9001:2016. *Systémy managementu kvality – Požadavky*. Praha: ÚNMZ, 2016.