

INFORMACE

Perspektivy důvěryhodného digitálního úložiště v rámci Národní digitální knihovny**Bohdana Stoklasová**

Národní knihovna ČR

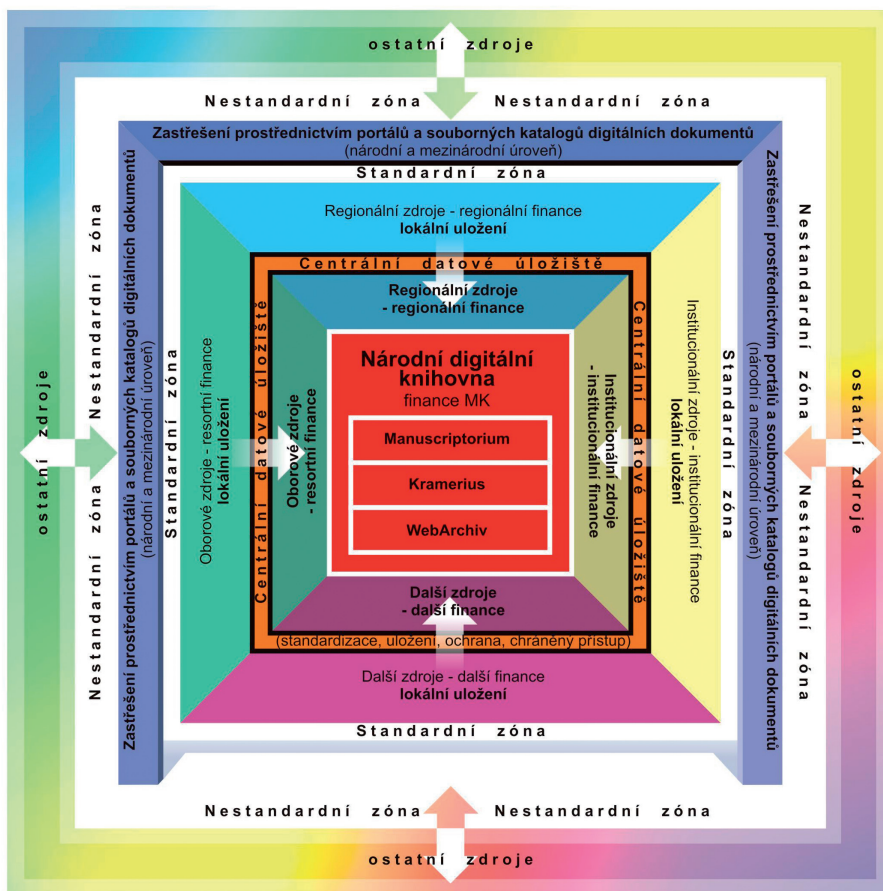
bohdana.stoklasova@nkp.cz

Problematika důvěryhodných digitálních úložišť je poměrně nová. Jedná se o oblast, která nepochybně v následujících letech projde v mezinárodním i národním kontextu bouřlivým vývojem a na stránkách periodika Knihovna se jí budeme detailněji věnovat. Následující příspěvek je stručným úvodem, který popisuje základní parametry důvěryhodného digitálního úložiště a snaží se rozptýlit mýtus o tom, že pro vybudování důvěryhodného digitálního úložiště stačí zakoupit potřebné technické a programové vybavení. V závěru se zamýšlíme nad tím, kdy a za jakých podmínek vznikne důvěryhodné digitální úložiště v rámci Národní digitální knihovny.

Důvěryhodné/á digitální úložiště v rámci Národní digitální knihovny a České digitální knihovny

Více či méně důvěryhodná digitální úložiště budou v budoucnosti vznikat jak v rámci Národní digitální knihovny, tak v širším kontextu České digitální knihovny. Jejich vzájemné propojení a delimitace odpovědností budou jedním z důležitých úkolů příštích let nejen pro knihovny, ale i pro další paměťové instituce. Postavení Národní digitální knihovny jako jádra České digitální knihovny i její vztah k ostatním subsystémům názorně vysvětlují grafické vyjádření a komentář k němu.

Česká digitální knihovna



Národní digitální knihovna (uprostřed) je tvořena vybraným souborem publikovaných digitálních i digitalizovaných dokumentů kvalifikovaných jako základ národního kulturního dědictví určeného k trvalému uchování a zpřístupnění současným i budoucím uživatelům. Digitální dokumenty tvořící jádro národního kulturního dědictví jsou v současnosti shromažďovány, uchovávány a zpřístupňovány v rámci tří národních programů: Memoria (Manuscriptorium – digitalizované historické dokumenty), Kramerius (digitalizované novodobé dokumenty) a WebArchiv (publikované digitální dokumenty). Jádro národního kulturního dědictví a důvěryhodné úložiště pro jeho trvalé uchování a zpřístupnění by mělo být financováno Ministerstvem kultury.

Širší Česká digitální knihovna je však tvořena velkým množstvím dalších digitálních dokumentů oborového, regionálního, institucionálního i jiného charakteru. Pouze některé z těchto zdrojů se kvalifikují jako nejcennější součást národního kulturního

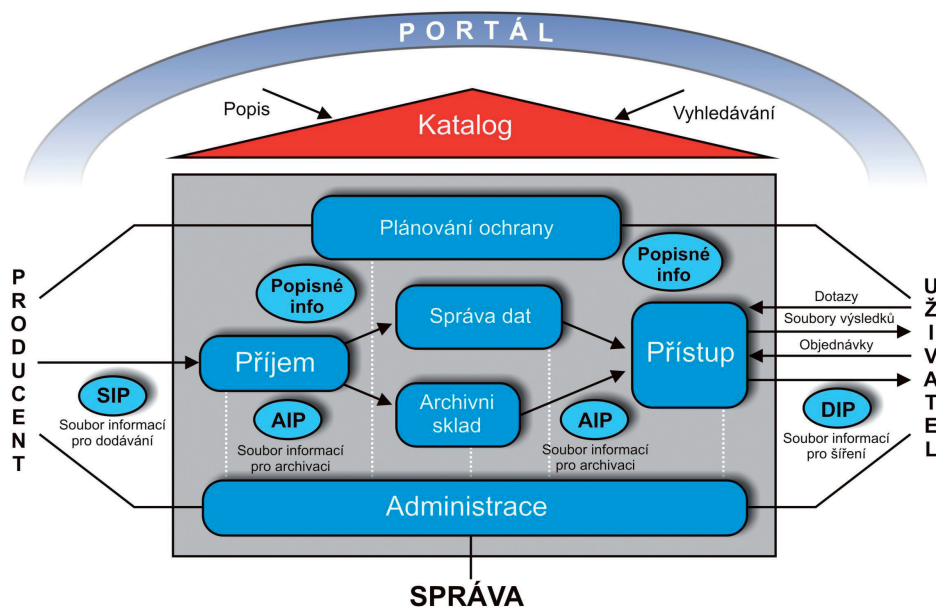
dědictví. Za shromažďování, trvalé uchovávání i zpřístupnění zdrojů mimo „jádro“ národního kulturního dědictví nesou odpovědnost (včetně finanční) příslušná resortní ministerstva, regiony, instituce atd.

Z obrázku je zřejmé, že kapacita důvěryhodného centrálního datového úložiště by měla v budoucnu umožnit uložit i dokumenty přesahující rámec jádra národního kulturního dědictví. Vzhledem k tomu, že vybudování i provoz důvěryhodného digitálního úložiště jsou finančně i personálně nesmírně náročné, řada institucí bude stát před otázkou, zda vybudovat a provozovat úložiště vlastní, nebo zda své cenné dokumenty svěžit (za úplatu) do úložiště centrálního.

Základní funkce důvěryhodného digitálního úložiště

Máme-li hledat odpověď na otázku, proč jsou vybudování a provoz důvěryhodného digitálního úložiště finančně i personálně natolik náročné, že překračují možnosti jednotlivých institucí, je třeba pochopit jeho základní funkce. V úplnosti je mapuje (a do detailů rozvíjí) referenční model OAIS (Open Archival Information System)¹, jehož rozšířená varianta (o katalog a portál) a analogie s funkcemi „klasické“ knihovny nám poslouží jako názorné pomůcky pro zjednodušený výklad základních funkcí digitálního úložiště.

Referenční model OAIS (Open Archival Information System) - ISO 14721



1 Reference Model for an Open Archival Information System (OAIS) : Blue Book. Consultative Committee for Space Data Systems, 2002.

Dostupný na: <http://www.ccsds.org/docu/dscgi/ds.py/Get/File-143/650x0b1.pdf>

Digitální úložiště (repozitář – repository) je definováno jako organizace, která uchovává informace s cílem jejich zpřístupnění a využití². Nejde tedy o prostý sklad digitálních informací, ale o komplex funkcí podobný funkcím klasické knihovny, která získává (přijímá) dokumenty od jejich producentů, zajišťuje jejich popis, uložení a správu ve skladišti a jejich zpřístupnění uživatelům, kteří si mohou informace o dokumentech vyhledat prostřednictvím katalogů nebo portálů a následně si je v knihovně půjčit. Digitální úložiště rovněž přijímá dokumenty (digitální) od jejich producentů, zpravidla včetně dohodnutých metadat (soubor informací pro dodávání). Výsledkem kontroly digitálních dokumentů a příslušných metadat a obohacení metadat (zde nehovoříme jen o metadatech popisných, ale i technických, strukturálních a administrativních)³ je soubor informací pro archivaci, při požadavku na zpřístupnění informací z úložiště je generován soubor informací pro šíření.

Zatímco u klasických dokumentů je relativně snadné zajistit jejich autenticitu a jejich ohrožení lze zjistit pouhým okem při prohlídce skladiště, u digitálních dokumentů je obojí podstatně složitější, stejně jako jejich zajištění proti neoprávněnému užití. Správa vlastních digitálních dokumentů i souvisejících metadat je složitý a permanentní proces, novou a prozatím málo prozkoumanou a zdokumentovanou oblastí je plánování ochrany digitálních dokumentů s ohledem na rychlý proces stárnutí a změn vlastních dokumentů (nutnost včasné migrace) i hardwaru a softwaru pro jejich zpřístupnění (nutnost včasné emulace v případě potřeby zachování nejen obsahu, ale i nosiče a celkového prostředí).

Zatímco degradace papírových nosičů informací je pomalá a snadno zjistitelná, ztráty v digitálním světě jsou rychlé, nevratné a ne vždy snadno a včas zachytitelné. Proto je v poslední době věnována v mezinárodním kontextu stále větší pozornost důvěryhodným digitálním úložištím.

Kritéria hodnocení důvěryhodnosti digitálního úložiště

Kritéria hodnocení důvěryhodnosti digitálního úložiště jsou obsažena ve dvou základních materiálech:

Trusted Digital Repositories : Attributes and Responsibilities (RLG, OCLC, 2002)² a *An Audit Checklist for the Certification of Trusted Digital Repositories : Draft for public comment* (RLG, NARA, 2005)⁴

2 *Trusted Digital Repositories : Attributes and Responsibilities*. Mountain View, CA : RLG, 2002. 62 s. Dostupný na: <http://www.rlg.org/legacy/longterm/repositories.pdf>

3 VOJNAR, Martin. Nové standardy digitálních knihoven pro dlouhodobou ochranu. In: *Knihovna*. Praha : Národní knihovna ČR, 2005, roč. 16, č. 2, s. 45–58. Dostupný na: <http://knihovna.nkp.cz/knihovna61/vojnar.htm>

4 *An Audit Checklist for the Certification of Trusted Digital Repositories : Draft for Public Comment*. Mountain View, CA : RLG, 2005. 65 s. Dostupný na: <http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf>

Starší z obou dokumentů *Trusted Digital Repositories : Attributes and Responsibilities*, vydaný v roce 2002, definuje obecné vlastnosti důvěryhodného úložiště takto:

Důvěryhodné úložiště musí:

- Přijmout odpovědnost za dlouhodobou péči o svěřené digitální dokumenty a za jejich zpřístupnění současným i budoucím uživatelům.
- Organizačně zajistit dlouhodobou životnost nejen pro vlastní úložiště, ale i pro svěřené digitální informace.
- Prokázat finanční zajištění v současnosti i trvale udržitelný rozvoj.
- Navrhnout systém pro správu digitálního úložiště v souladu s obecně platnými konvencemi a standardy v zájmu zaručení trvalé správy, zpřístupnění a zabezpečení uložených digitálních dokumentů.
- Stanovit metodiku hodnocení důvěryhodnosti systému.
- Jasně a srozumitelně prezentovat svoji odpovědnost za dlouhodobou ochranu a zpřístupnění dokumentů uživatelům i subjektům, které své dokumenty v úložišti deponují.
- Disponovat strategií, pracovními postupy a službami, které umožňují snadné hodnocení a měření.

Navazující dokument *An Audit Checklist for the Certification of Trusted Digital Repositories* již představuje poměrně detailní a propracovaný systém kritérií pro hodnocení důvěryhodného úložiště.

Pro účely tohoto obecného úvodního příspěvku se budeme pohybovat jen na dvou nejvyšších úrovních, v následujících příspěvcích uveřejňovaných postupně v tomto periodiku se budeme věnovat jednotlivým okruhům podrobněji.

Hodnocení spolehlivého digitálního úložiště spadá do čtyř hlavních sekcí:

- A. Organizace**
- B. Funkce, postupy, metody**
- C. Cílová skupina uživatelů a využitelnost informací**
- D. Technologie, technická infrastruktura**

V jednotlivých sekcích jsou sledovány okruhy, v jejichž rámci je třeba zodpovědět poměrně konkrétní otázky. Prozatím zůstáváme na úrovni obecných okruhů, jejichž obsah je ve většině případů jasný z jejich názvu. Pokud název není dostatečně zřetelný, jsou připojeny vysvětlivky v závorce. Další podrobnosti je možné nalézt přímo v dokumentu.

A. Organizace

- A.1. Řízení organizace a schopnost „přežítí“ (dlouhodobá ochrana a zpřístupnění jako strategické priority, schopnost přestát krizi bez výrazných ztrát)
- A.2 Aдекватní organizační struktura a personální zajištění
- A.3 Jasná a srozumitelná dokumentace všech procedur, možnost snadného vyhodnocení

- A.4 Finanční stabilita
- A.5 Kontrakty, licence, odpovědnost

B. Funkce, postupy, metody

- B.1 Příjem/získávání obsahu
- B.2 Archivace a správa uložených informací
- B.3 Plánování ochrany: migrace a další strategie
- B.4 Správa dat
- B.5 Správa přístupu

C. Cílová skupina uživatelů a využitelnost informací

- C.1 Jasně vymezená a zdokumentovaná cílová skupina uživatelů (její potřeby, znalosti – jazykové, technické atd.)
- C.2 Popisná metadata odpovídající potřebám cílové skupiny uživatelů
- C.3 Způsob a možnosti využití (obsah, doba odezvy, licenční podmínky a smlouvy, dostupnost dat, kontrola dodržování podmínek)
- C.4 Srozumitelnost výstupů pro cílovou skupinu uživatelů (ověřování)

D. Technologie, technická infrastruktura

- D.1 Systém – spolehlivost, charakter systému adekvátní významu úložiště
- D.2 Technologie – HW, SW
- D.3 Zabezpečení, plánování ochrany, krizový plán (disaster & recovery)

Z uvedeného výčtu je zřejmé, že pro vybudování a provoz důvěryhodného úložiště zdaleka nestačí zakoupení drahého technického a programového vybavení. Jedná se o složitý a dlouhodobý proces, který musí být řádně zakotven ve strategických prioritách i organizační struktuře instituce, která aspiruje na vybudování a provoz důvěryhodného digitálního úložiště. Musí být adekvátně finančně a personálně zajištěn nejen pro dobu vzniku úložiště a po krátkou dobu po něm, ale dlouhodobě. Tvorba a provoz spolehlivých digitálních úložišť jsou po všech stránkách natolik naléhavé a zároveň náročné, že přesahují možnosti i těch největších institucí. Proto se jedná o velkou výzvu ke spolupráci paměťových institucí v národním i mezinárodním kontextu.

Perspektivy a harmonogram vytvoření důvěryhodného digitálního úložiště v rámci Národní digitální knihovny

Vybudování důvěryhodného digitálního úložiště je součástí Koncepce trvalého uchování knihovních sbírek tradičních a elektronických dokumentů v knihovnách ČR. Etapa1: do roku 2010⁵. Koncepce čeká na Ministerstvu kultury ČR na předložení k projednání a schválení ve vládě, na němž závisí i její finanční zajištění, bez kterého jsou vybudování i provoz spolehlivého digitálního úložiště nereálné.

⁵ Koncepce trvalého uchování knihovních sbírek tradičních a elektronických dokumentů v knihovnách ČR. Etapa1: do roku 2010. Praha : Národní knihovna ČR, 2006
Dostupný na: <http://www.ndk.cz/koncepce>

V roce 2006 proběhla analýza funkcí důvěryhodného digitálního úložiště a byla jmenována Pracovní skupina pro koordinaci tvorby a využívání českých digitálních úložišť <http://www.ndk.cz/pracovni-skupina/clenove-pracovni-skupiny/> . Někteří členové pracovní skupiny navštívili fungující evropská i mimoevropská důvěryhodná digitální úložiště. Národní knihovna ČR provedla analýzu stavu dlouhodobého uchování a zpřístupnění digitálních dokumentů v evropských národních knihovnách⁶. Z analýzy vyplynulo, že pro většinu evropských národních knihoven je tato oblast jednou z hlavních současných strategických priorit. Existence důvěryhodných digitálních úložišť v Evropě je (na rozdíl od ostatních světadílů) sice zatím výjimkou, do roku 2010 však lze očekávat prudký rozvoj v této oblasti.

V roce 2007 bude pokračovat detailnější analýza funkcí důvěryhodného digitálního úložiště včetně návštěv dalších institucí. Bude probíhat analýza metadat (popisných, technických, strukturálních i analytických) a příprava žádosti o nabídky (RFP) pro komplexní systém pro správu úložiště. Stav v závěru roku 2007 by měl zhodnotit interní audit.

V roce 2008 bude dokončena žádost o nabídky (RFP) pro komplexní systém pro správu úložiště a proběhne výběrové řízení a instalace systému. V roce 2009 bude systém fungovat v poloprovozním režimu, na počátku roku 2010 bude zahájen plný provoz a proběhne mezinárodní audit.

Pokud dojde ke schválení a finančnímu zabezpečení Koncepce trvalého uchování knihovných sbírek tradičních a elektronických dokumentů v knihovnách ČR. Etapa1: do roku 2010, bude mít Národní knihovna důvěryhodné, mezinárodně certifikované digitální úložiště v roce 2010.

6 Dlouhodobé uchování a zpřístupnění digitálních dokumentů v Evropě : výsledky dotazníkového průzkumu. Praha : Národní knihovna ČR, 2006
Dostupný na: <http://www.ndk.cz/dokumenty/dpedotaznikcz.doc/view>

