



**ÚSTAV INFORMAČNÍCH STUDIÍ A KNIHOVNICTVÍ
FF UK V PRAZE**

Jan Hutař

Úvod do ochrany digitálních dat

Verze 1.0

Praha

Listopad 2008

1 ÚVOD

Význam digitálních informací je v dnešní době těžko přehlédnutelný a setkáváme se s ním všichni v různých oborech. Digitální prostředí a informace se již dávno rozšířily do oblastí akademických, průmyslových, vládní administrativy a samozřejmě také do domácností. Velké množství informací v digitální podobě jako např. text, obrazové informace, audio, video, webové stránky, počítačové programy, databáze, je vytvářena, sdílána a využívána za různých podmínek pro nesmírné množství účelů. Toto rozličné využití digitální technologie spočívá na výhodách s ní spojených, např. na možnostech vyhledávání, sdílení, možnosti perfektní duplikace a také na interoperabilitě*.

Růst výpočetní síly a propustnosti sítí vede k neustálému nárůstu využití digitálních informací. S touto skutečností se pojí další, a to ta, že mnoho organizací, které těchto digitálních informací již delší dobu využívají, se nyní dostává do situace, kdy musí chtít nechtě přistoupit k nějakým „archivním“ opatřením. Ochránit a zachovat digitální data v dlouhodobém horizontu je ovšem daleko těžší než např. v případě papíru. Je to dáno i tím, že hardware a software se stávají zastaralými během let (někdy i měsíců)[†].

Nejvíce patrné je to na oblasti *digital preservation (ochrany digitálních objektů[‡]) - zajištění dlouhodobé životnosti informací v digitální formě*. Možnosti a schopnosti vytvářet a využívat digitální informace se neustále rozvíjejí, ovšem zajistit dlouhodobou ochranu těchto informací bylo a je velmi zdoluhavé a náročné. Tento problém začal být bolestivým místem před relativně krátkou dobou, kdy se ochrana digitálních materiálů stala naléhavou a závaznou, jako následek křehkosti úložných digitálních médií, jakož i rapidního zastarávání úložných a zobrazovacích prostředí. Všudypřítomnost otázek ohledně digital preservation zatím vytváří obecné základy pro mezioborový dialog a spolupráci pro lepší odpovědi na všechny výzvy tohoto problému.

Pro zajímavost uvedu jednu z mnoha definic „digital preservation“. Vznikla v roce 2007 v rámci pracovní skupiny Americké asociace knihoven, která publikovala tři definice, krátkou, středně dlouhou a dlouhou. Střední definice[§]:

„Digital preservation spočívá v kombinaci plánů, strategií a opatření pro zajištění přístupu k reformátovanému a digitálně vzniklému digitálnímu obsahu bez ohledu na problémy spojené s nestálostí médií a technologickými změnami. Cílem digital preservation je přesné zobrazení autentického obsahu v jakémkoliv časovém horizontu od jeho vzniku.“

Podobnou definicí může být tato od H.M. Gladneyho^{**}, který říká, že digital preservation je „organizované opatření k zajištění dlouhodobé použitelnosti digitálních objektů; zásadní je že digitální objekty nebudou nikdy ztraceny nebo poškozeny, jsou důvěryhodné, je možné je vždy najít, rozumět jim a to i přes problémy zastarávání technologií.“

Naléhavost a potřeba k podniknutí prvních kroků ochranných opatření k zajištění dlouhodobé životnosti digitálního materiálu zasáhla všechny instituce (podniky, média, vládní úřady apod.) i jednotlivce (např. fotografové). Mnoho významných organizací si neuvědomuje význam archivní funkce v rámci svých mandátů. Velmi pozitivní je, že jakékoliv úsilí přijít s nějakým řešením v rámci jedné komunity, má velmi často efekt mexické vlny na komunity z relativně jiné oblasti^{††}.

Právě takovou aktivní komunitou a oblastí kde vzniká velké množství digitálních dat a kde se díky tomu dostává ke slovu ochrana digitálních dat, je oblast tzv. paměťových institucí. Knihovny, muzea i archivy vytvářejí velké množství digitálních objektů (dokumentů) v nákladných digitalizačních projektech a uvědomují si, že takto vzniklé digitální objekty je nutné bezpečně uložit, ochraňovat je a zajistit jejich dostupnost i v dlouhodobém časovém horizontu.

* LAVOIE, Brian F. *The Open Archival Information System Reference Model*. s. 1.

† STOKLASOVÁ, Bohdana; HUTAŘ, Jan. *Nové směry v dlouhodobém uchování dokumentů...* s. 87.

‡ budu v dokumentu používat dále anglický výraz Digital preservation, pro jeho jednoznačnost a jednoduchost

§ American Library Association. *Definitions of Digital Preservation*

** GLADNEY, Henry M. *Preserving digital information*. Berlin : Springer, 2007. s. 270

†† LAVOIE, Brian F. *The Open Archival Information System Reference Model*. s. 1.

2 DIGITÁLNÍ REPOZITÁŘ

Digitální repozitář je v podstatě informační systém. Má všechny části/podsystemy, které takový informační systém vytvářejí:

- systém vytváření informací (dodavatel, tvůrce)
- systém zprostředkování informací (repozitář samotný)
- a systém využívání informací (uživatel, jiný repozitář, instituce samotná)

Na digitální repozitář se lze také dívat jako na informační pramen, kterým bezpochyby také je. Dokumenty v něm jsou elektronické/digitální a v hmotné podobě neexistují, dají se přenášet, po síti nebo pokud se uloží na nějaké paměťové médium, a dají se zobrazovat, což je asi jediná forma zpřístupnění uživateli.

2.1 DEFINICE DIGITÁLNÍHO REPOZITÁŘE A JEHO TYPY

Obecně řečeno, digitální repozitář má za cíl sbírat, uchovávat (krátkodobě nebo i dlouhodobě) a zpřístupňovat informace, které jsou do něj ukládány, ať již automaticky nebo ručně. S obecnou definicí digitálního repozitáře to není tak jednoznačné, repozitářů existuje několik druhů, které se liší nejen svou politikou (strategií), ale hlavním důvodem odlišnosti je, že závisí na potřebách organizace, která je provozuje. Takovou organizací může být např. univerzita, knihovna, archiv, výzkumný ústav apod.

Pro ilustraci, cílem určitého repozitáře může být uchovávat informace a poskytovat je ve srozumitelné podobě uživatelům, zatímco cílem jiného repozitáře je také uchovávat informace, ke kterým ale neposkytuje vůbec žádný přístup (někdy je označován jako „dark archive“). Lze tedy říci, že určující **společnou vlastností jakéhokoliv repozitáře je informace uložit, tedy uchovat**, což je jeho nejpodstatnější úkol.

Jako taková obecná definice by se tedy dala použít např. tato: *“místo na které lze uložit různé digitální materiály, a které má přidaný smysl v tom co ukládá a k jakému účelu je provozováno*.”*

Je velmi důležité si uvědomit, že digitální repozitář není jen několik kusů hardwaru, který spolu funguje podle nějakého zadání, ale že nedílnou součástí digitálního repozitáře je i organizace, která ho spravuje, se všemi svými slabými a silnými stránkami. Od této organizace se odvíjí to, jak repozitář funguje, v jakých podmínkách apod. Repozitář je tak spolehlivý, jak silná a spolehlivá je organizace, která repozitář pro své účely provozuje.

Repozitář může mít několik následujících charakteristik:

- v repozitáři je uložen nějaký obsah – může být uložen buď autorem toho obsahu, nebo vlastníkem nebo v jejich jméně
- architektura repozitáře zahrnuje správu jak obsahu tak i metadat
- repozitář poskytuje alespoň základní služby (uložit, vyhledat, poskytovat dokumenty)
- repozitář musí být udržitelný a důvěryhodný ve smyslu, že je velmi dobře podporován a spravován (viz dále v textu)

2.2 REFERENČNÍ MODEL OAIS - OPEN ARCHIVAL INFORMATION SYSTEM

Úsilí vyvinout archivní standardy pro dlouhodobé ukládání dat vzniklo na základě potřeby vývoje datového standardu pro podporu vesmírného výzkumu a uložení vědeckých dat. Bylo zadáno jako potřeba International Organization for Standardization (ISO), konkrétněji Consultative Committee for Space Data Systems (CCSDS), což je orgán pro mezinárodní spolupráci vesmírných agentur. Na začátku této snahy navrhlo CCSDS referenční model, který měl ukotvit terminologii a koncepty pro popis a porovnání datových modelů a archivních architektur; popsat podstatné entity a

* CARPENTER, Leona. *Repositories in Context*. UKOLN, 2005.

vztahy mezi nimi v archivním prostředí; vysvětlit klíčové funkce a informační komponenty archivního systému a nakonec, měl posloužit jako rámec pro další aktivitu. Vše se povedlo, včetně podpory dalších aktivit.

Práce CCSDS vyústila v květnu 1999 ve vydání referenčního modelu OAIS. Referenční model je koncepční rámec pro jakýkoliv archivní systém věnovaný problému ochrany a správy přístupu k digitálním informacím v dlouhodobém horizontu. Popisuje prostředí, ve kterém digitální archiv (repozitář) sídlí, jeho funkční komponenty a informační infrastrukturu podporující všechny procesy archivu. Referenční model podstoupil review jako doporučení pro normu ISO, ale hlavně, rozšířil se daleko mimo komunitu vesmírných výzkumů do knihoven a ostatních paměťových, vládních institucí i do soukromého sektoru. Upravená verze modelu byla publikována v r. 2001*.

Referenční model OAIS se ukázal jako velmi životaschopný a je dnes široce implementován a využíván. Je ideální svou obecností a tím tedy i možností implementace. Celý rámec v podstatě popisuje entity a procesy, které by měl repozitář zahrnovat a dělat, s tím, že určité části tohoto rámce mohou ve skutečném repozitáři chybět.

Je ovšem nutné si uvědomit, že referenční model OAIS představuje obecný (high level) popis typů informací vytvářených a spravovaných funkčními komponenty celého archivního systému a *nezabývá se konkrétními typy digitálních objektů, které se spravují v archivu, ani specifikací technologie[†] použité v archivu, aby se dosáhlo cíle, tj. ochrany a údržby přístupu k digitálním objektům v dlouhodobém horizontu.* Model jako takový je použitelný pro všechny datové archivy a ty organizace nakládající s informacemi, které mohou potřebovat dlouhodobou ochranu.

Referenční model odráží zaměření na digitální informace, ať již na primární podobu archivované informace a také na informace podpůrné pro digitální nebo fyzické archivované materiály. Nicméně model lze aplikovat in na ne-digitální informace (např. fyzické předměty), ovšem ochrana a architektura takových informací není rozebrána do detailu. Referenční model mj.[‡]:

- poskytuje rámec pro pochopení a zvýšení povědomí o archivních konceptech potřebných pro dlouhodobou ochranu digitálních informací a přístup k nim
- poskytuje koncept potřebný pro nearchivní instituce k tomu, aby byly užitečnými účastníky v procesech ochrany
- poskytuje rámec, včetně terminologie a konceptů, pro popis a porovnání architektur a procesů již existujících i budoucích archivů
- poskytuje rámec pro popis a porovnání různých strategií a technik dlouhodobé ochrany
- poskytuje základ pro porovnání datových modelů digitálních informací ochraňovaných v archivech a také pro diskuzi o tom, jak se mohou datové modely a zásadní informace během doby měnit
- poskytuje základ, který může být rozšířen dalším úsilím tak, aby pokryl dlouhodobou ochranu informací, které nejsou v digitální formě (např. fyzická média a fyzické vzorky).
- dá základ rozpoznání a vytváření s OAIS souvisejících standardům

2.2.1 Referenční model OAIS a OAIS archiv (repozitář)

Referenční model se zabývá celou škálou ochranných funkcí archivních informací včetně sběru (ingest), archivního uložení, data managementu, přístupu a rozšiřování informací. Zabývá se rovněž migrací digitálních informací na nová média a formy, data modelem použitým na reprezentaci informací, rolí softwaru v ochraně informací a výměnou digitálních informací mezi archivy. Určuje zároveň jak interní, tak externí pracovní prostředí archivních funkcí a určuje také mnoho obecných služeb těchto prostředí. Poskytuje také různé příklady a best practice doporučení. Definuje také minimální sadu odpovědností pro archiv/repozitář, aby mohl být nazván *OAIS archivem*.

**Preservation Metadata and the OAIS Information Model*. s. 5.

† tj. není to návod na implementaci určitých technologií

‡ CCSDS 650.0-B-1: *Reference Model for an Open Archival Information System (OAIS)*.

OAIS je archiv sestávající z organizace lidí a systémů, kteří přijali odpovědnost za ochranu informací a jejich zpřístupnění určité komunitě. To představuje určité odpovědnosti, definované v rámci OAIS, a které dovolují takový OAIS archiv odlišovat od ostatních systémů popisovaných slovem „archiv“.

Hlavní koncept referenčního modelu vychází právě ze zkratky OAIS. Výraz open odkazuje na skutečnost, že referenční model byl vytvořen a volně publikován na veřejných fórech, na kterých mohl kdokoliv zainteresovaný participovat. Pozor - slovo nenaznačuje, že přístup do archivu je neomezený. Archival information system je „organizace lidí a systémů (hardware, software), kteří vzali za svou odpovědnost ochránit informace a zpřístupnit je pro určitou komunitu.“

Tato definice zdůrazňuje dvě primární funkce archivních repozitářů: za prvé **ochránit informaci**, tj. zajistit její dlouhodobou odolnost, za druhé **poskytnout přístup** k archivované informaci, v závislosti na potřebách uživatelů nebo komunity. Tato definice je podpořena výčtem nároků, které musí archiv OAIS typu splňovat. OAIS musí obzvláště*:

- vyjednávat a získávat informace od producentů informací.
- zajistit dostatečnou kontrolu informací aby se splnily podmínky dlouhodobé ochrany.
- vymežit hlediska využití uživatelskou komunitou.
- zajistit aby byly chráněné informace nezávisle srozumitelné uživatelům v tom smyslu, že je možné informaci rozumět a vnímat bez využití asistence producenta této informace.
- dodržovat strategii a procedury a tím zajistit to, že informace je ochráněna proti všem nepředvídatelnostem, je možné rozšiřovat autentické kopie informace v originální formě nebo ve formě, která na tu původní odkazuje.
- zpřístupnit chráněné informace uživatelské komunitě.

Spravované informace jsou považovány za hodné k dlouhodobému uchování, jakkoliv samo OAIS není permanentní. Termín „dlouhodobý“ označuje časový úsek dost dlouhý na to, abychom uvažovali o jeho dopadu na změny technologií, podpory formátů nových i starých, nových médií a také o změnách cílové komunity. „Dlouhodobost“ se může nepředvídatelně prodlužovat.

2.2.2 OAIS – informační balíčky – Information Package (IP)

Repozitář odpovídající OAIS pracuje s několika typy tzv. informačních balíčků, které model OAIS definuje. Ty mohou ideálně obsahovat informace o svém obsahu a samozřejmě k nim informace podpůrné. Jsou to SIP, AIP a DIP – podrobněji viz níže v textu této části.

Informace o obsahu

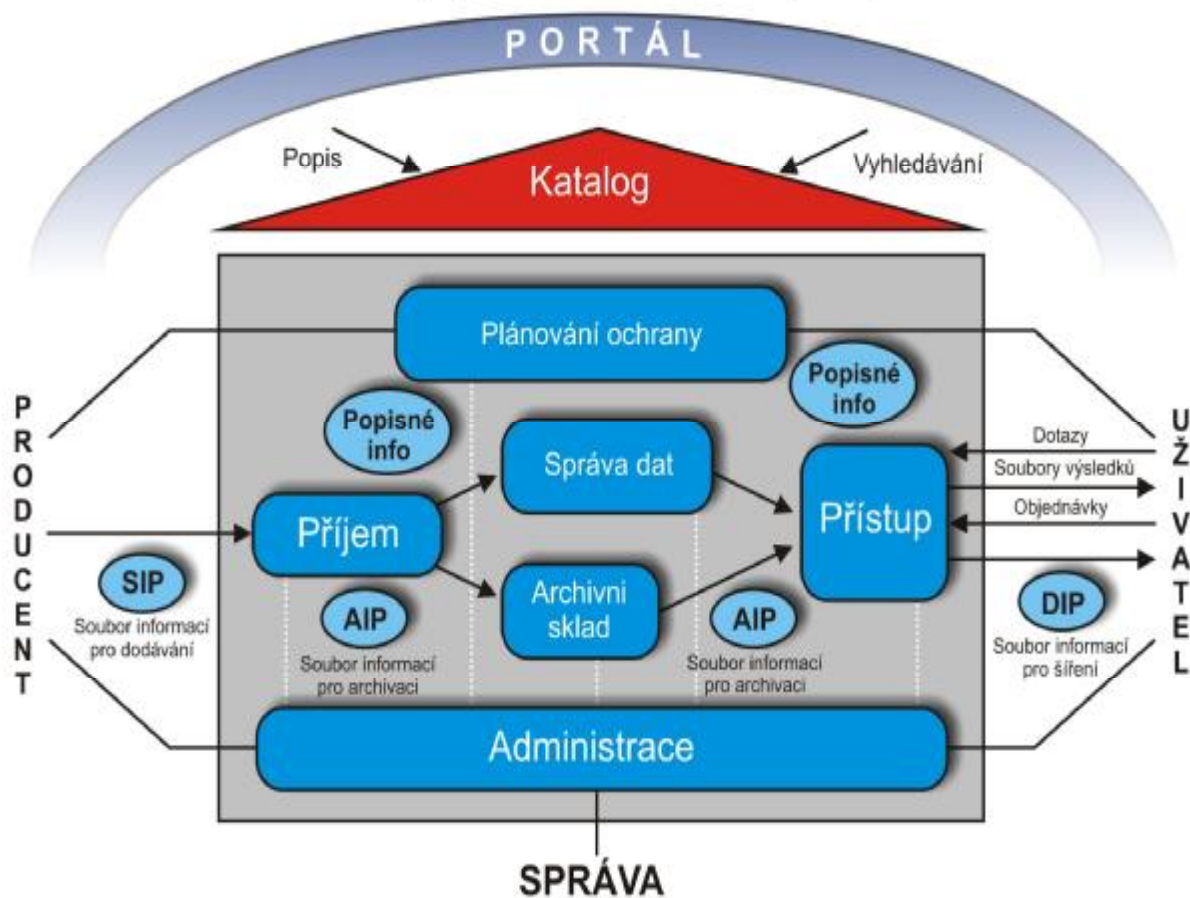
- je sama o sobě srozumitelná určité komunitě, tj. informace je v podobě souboru.
- může být i fyzickým objektem.
- potřebuje Reprezentaci informace, aby dala smysl sledu bitů v souboru.

Podpůrné informace, tj. to

- jak vlastně vznikla Informace o obsahu.
- jak souvisí s ostatními informacemi.
- jak je zajištěna její integrita.
- taktéž má tzv. Reprezentaci informace (viz kapitola 2.3).

* LAVOIE, Brian F. *The Open Archival Information System Reference Model*. s. 3.

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



Submission Information Package - SIP

- je předmětem vyjednávání mezi producentem (tvůrcem informací) a OAIS/repozitářem
- je zaslán producentem (dodavatelem) do OAIS/repozitáře

Archival Information Package - AIP

- balíček informací používaný pro vlastní archivaci digitálních objektů
- obsahuje kompletní sadu Popisné ochranné informace (Preservation Description Information - PDI) pro obsahové informace

Dissemination Information Package - DIP

- obsahuje část nebo všechny části jednoho nebo více AIP
- OAIS/repozitář je rozšiřuje uživatelům z konkrétní komunity (tj. je to vlastně to, co se dostane z repositáře uživatelům)

2.2.3 Shrnutí aneb co je vlastně referenční model*?

Rámec

- pro pochopení podstatných vztahů mezi entitami určitého prostředí a
- pro vývoj důsledných standardů nebo specifikací toto prostředí podporujících.
- pro pochopení a aplikaci konceptu potřebného pro dlouhodobou ochranu digitálních informací.
- pro pochopení a implementaci konceptů potřebných pro dlouhodobou ochranu digitálních informací.

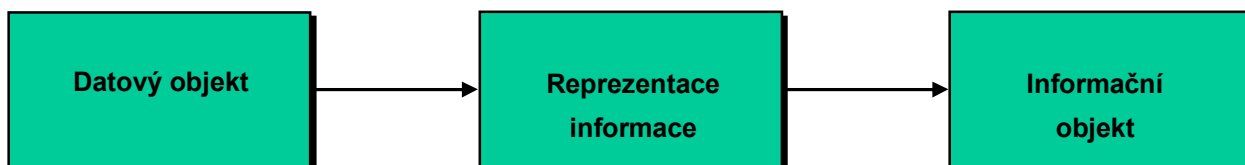
Referenční model

- je založen na malém počtu jednotných konceptů.
- je abstrakcí klíčových konceptů, jejich vztahů a jejich prostředí jak mezi sebou tak vůči okolnímu prostředí.
- může být využit jako základ pro vzdělávání a vysvětlování standardů pro laiky (ne-specialisty).

Zajímavostí je, že rámec do jisté míry odpovídá koloběhu, který se provádí s klasickými dokumenty v knihovnách nebo archivech. Je tady jistá analogie mezi uložením a zpřístupněním klasických dokumentů a dokumentů digitálních. Digitální je ovšem složitější a mnohem zranitelnější.

2.3 POJETÍ INFORMACE V REFERENČNÍM RÁMCI OAIS‡

Informace je vždy vyjádřena (tj. reprezentována) pomocí nějakého typu dat. Data interpretovaná pomocí své „Reprezentace informace“ dávají vzniknout informacím (informačnímu objektu). Ochrana „Informačního objektu“ vyžaduje jasnou identifikaci a pochopení datového objektu a s ním spojené „Reprezentace informace“.



Referenční model OAIS je znázorněn na obrázku níže. V kontextu OAIS může informace existovat ve dvou formách – buď jako fyzický objekt (tj. např. papírový dokument, vzorek půdy), nebo jako digitální objekt (např. PDF soubor, TIFF soubor). Tyto dva druhy – fyzický a digitální – jsou obecně známy jako tzv. Datové objekty. Takový datový objekt může nabývat různých forem, ať je to materiál, který je cílem ochrany, nebo metadata asociovaná s archivním digitálním objektem[§].

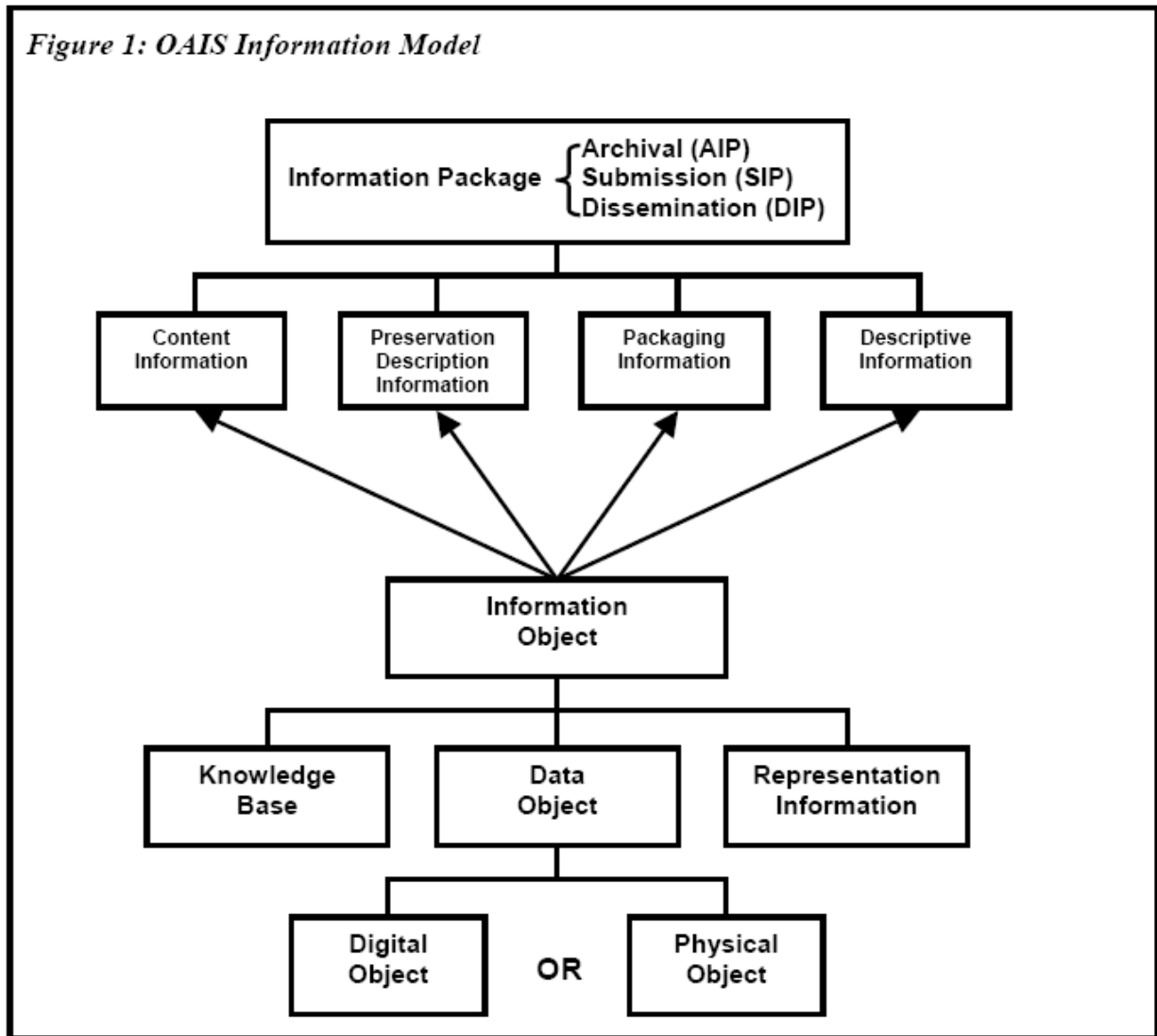
* SAWYER, Don; REICH, Lou. *ISO "Reference Model For an Open Archival Information System (OAIS)"*.

† STOKLASOVÁ, Bohdana. *Dlouhodobé uchovávání dokumentů v českých knihovnách [online]*.

‡ SAWYER, Don; REICH, Lou. *ISO "Reference Model For an Open Archival Information System (OAIS)"*.

§ *Preservation Metadata and the OAIS Information Model*. s. 5.

Figure 1: OAIS Information Model



2.4 K PŘIPOMENUTÍ*

- OAIS je pouze konceptuální model a NE model architektury konkrétního systému.
- Funkce popsaného OAIS referenčního modelu nemusí nutně odpovídat funkčním modulům systému, který má odpovídat modelu OAIS.
- Funkčnost konkrétního digitálního repozitáře může být daleko více omezená, než je vyjádřeno v referenčním modelu OAIS.

* SAWYER, Don; REICH, Lou. *ISO "Reference Model For an Open Archival Information System (OAIS)"*.

3 CERTIFIKOVANÝ DŮVĚRYHODNÝ DIGITÁLNÍ REPOZITÁŘ

S pokračujícím vývojem a již běžným využíváním repozitářů se začalo mluvit o tzv. důvěryhodných repozitářích. Co vlastně má označovat slovo „důvěryhodný“? Označuje schopnost repozitáře zachovat digitální dokumenty v **dlouhodobém horizontu** přístupné a použitelné. Tedy repozitář musí být navenek důvěryhodný, což se netýká pouze technického řešení, právě naopak. Jde o to, jak je zajištěna organizace, ve které je repozitář provozován (např. financování, personálně odborné zabezpečení, status organizace, mandát apod.).

V rámci repozitáře se většinou počítá s tím, že informace na něm budou uloženy velmi dlouho, přičemž budou stále v použitelné, tedy čitelné podobě, která se dá nabídnout uživatelům. Tohle platí hlavně pro tzv. certifikované důvěryhodné repozitáře. Vyhovět jednomu z hlavních, ne-li hlavnímu cíli certifikovaných důvěryhodných repozitářů není v době překotného vývoje technologií vůbec jednoduché. Technologie se vyvíjí obrovským tempem, což přináší nejen zastarávání hardwaru, ale hlavně zastarávání softwaru. Je skoro jisté, že formáty jakýchkoliv digitálních dokumentů (text, obraz, video, audio), které jsou dnes rozšířeny a obecně využívány (.doc, .pdf, .mp3, .avi, .jpg apod.), budou v blízké budoucnosti nahrazeny jinými, lepšími a později bude mít celá společnost seriózní problém digitální dokumenty ve starých formátech vůbec použít (přečíst, přehrát, upravit nebo i otevřít).

Nemluvíme zde o něčem, co se teprve stane, ale o něčem, co se již dlouhou dobu děje. K takové markantní proměně softwaru již došlo i například v oblasti textových editorů. Kdo si dnes vzpomene na textový editor AmiPro, běžící na počítačích Amiga ještě před nástupem Microsoft Office a jeho Wordu? To samé se týká obrazových digitálních dokumentů, které procházejí ještě bouřlivějším vývojem, který žene kupředu snaha o co největší komprimaci při co nejlepším zachování kvality.

V tomto smyslu se na problém zastarávání softwaru a hardwaru můžeme dívat jako na jednu z forem technické informační bariéry, která ohrožuje velkou část našich digitálních dějin. Víme, že úroveň celé společnosti závisí na úrovni a dostupnosti informací[†].

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 hardware a software pro jejich zpřístupnění (nutnost včasné emulace v případě potřeby zachování nejen obsahu, ale i nosiče a celkového prostředí). Degradace papírových nosičů informací je pomalá a snadno zjiřitelná, ztráty v digitálním světě jsou naopak rychlé, nevratné a ne vždy snadno a včas zachytitelné. Proto je v poslední době v mezinárodním kontextu věnována stále větší pozornost důvěryhodným digitálním repozitářům[‡].

3.1 PROBLÉMY S DLOUHODOBOU ARCHIVACÍ DIGITÁLNÍCH OBJEKTŮŠ

Jaké druhy problémů vlastně mohou nastat v případě, že se digitální objekty budou uchovávat např. několik desítek nebo i set let?

* Pokud bychom chtěli jít do extrému v oblasti hardwaru, můžeme za takovou změnu považovat náhradu děrných štítků za magnetické pásky, posléze za pevné paměti, optické disky apod. Co si kdo dnes počne s děrnými štítky, kdyby z nich chtěl interpretovat nějakou informaci? Nemusíme ani chodit tak daleko do minulosti, problémem se mohou přeci ukázat i obyčejné diskety, které jsme před několika lety používali denně, a které jsou dnes vytlačeny flashovými paměťmi a na prodávaných noteboocích a počítačích disketová mechanika zcela chybí.

† KÖNIGOVÁ, Marie. *Vybrané kapitoly z informační vědy*. s. 17

‡ STOKLASOVÁ, Bohdana; HUTAŘ, Jan. *Nové směry v dlouhodobém uchovávání dokumentů...* s. 87.

§ GIARETTA, David. *ISO/CCSDS Open Archival Information System (OAIS) ...*

1. Archivovaná informace musí být použitelná uživateli, kteří jsou vzdáleni v čase, prostoru a nemají podporu producenta té informace.
 - Může se stát, že producent informace již neexistuje, v tom případě nemůže odpovědět na žádné otázky. Jediné co máme, je digitální objekt a jeho popis (metadata).
 - Software na kterém byla informace (digitální objekt) vytvořena již nemusí být podporován. Může se tedy stát, že informace zaznamenané tímto softwarem jsou zcela nedostupné (nikdo např. nezná kódování pro jejich využití a dokumentace se během let ztratila). Uživatel nemusí být schopen ani rozeznat, zda digitální objekt je text, obraz či zvuk.
2. Uživatelská komunita se bude během doby měnit.
 - Nová komunita nebude znát pozadí konkrétních informací (vznik, software, hardware, účel ...)
 - Nová komunita může používat naprosto odlišné pracovní prostředí.
 - Může chtít kombinovat informace z mnoha zdrojů.
3. Repozitář se bude měnit během doby.
 - Může dojít k migracím na nové technologie, což vždy souvisí s reorganizací informací v repozitáři a ta zase s možnými změnami ve vztazích mezi informacemi.

Může také dojít k migraci do jiných institucí, tj. možné změny v managementu, datové struktuře i formátu souborů.

3.2 EMULACE A MIGRACE

V rámci (nejen) certifikovaných důvěryhodných repozitářů by systém repozitáře sám od sebe měl poznat, kdy který formát digitálních objektů zastarává a provést např. jeho emulaci nebo migraci. Samozřejmě to systém neudělá sám od sebe, ale na základě určitých informací (metadat) a na základě spolupráce s externími registry formátů (např. PRONOM^{*} nebo GDFR[†]). Digitální dokumenty obsahují maximální množství metadat již při svém vstupu do repozitáře (tj. poskytně je původce digitálního objektu). Existují i nástroje, které automaticky z digitálních objektů při jejich vstupu do repozitáře potřebná metadata vytvoří (většinou technická metadata). Jedním z nástrojů je např. známý JHOVE[‡].

Pod pojmem *emulace* se rozumí umělé „napodobení“ softwarového a hardwarového prostředí, které bylo typické nebo nutné pro prohlížení digitálního objektu v době jeho vzniku. Příkladem může být emulační software, který dokáže napodobit hardwarové vybavení počítače z roku 1995 a zároveň v rámci tohoto emulátoru spustit software z té doby (např. MS Windows 95). Nepopíratelnou výhodou tohoto přístupu je, že můžeme digitální objekt v repozitáři uložený v jeho původní podobě a nemusíme ho jednou za čas transformovat do stále nových a nových formátů. Nevýhodou naopak je, že emulační software stále není vyvinut natolik, aby tento přístup poskytoval záruku, že digitální objekty opravdu bude možné kdykoliv zpřístupnit v jejich původním prostředí. Navíc emulační software čelí stejnému problému, v horizontu několika let bude sám potřebovat další emulátor, aby bylo možné původní emulátor spustit. Příkladem emulačního softwaru pro potřeby digital preservation je např. DIOSCURI[§] vyvinutý ve spolupráci nizozemské národní knihovny a národního archivu.

Migrace je druhou metodou aplikovanou v rámci ochrany digitálních objektů. Je to přístup, který reflektuje neustálý vývoj formátů a migruje ve vhodný okamžik všechna data v konkrétním formátu uložená v repozitáři do formátu novějšího, který je v tom okamžiku aktuální. Cílem je zamezit tomu, že budou v repozitáři uložena data ve formátu, který už nemá podporu aktuálních softwarových aplikací a není tedy možné objekt zobrazit. Příkladem budiž převod textového

* <http://www.nationalarchives.gov.uk/pronom/>

† <http://www.gdfr.info/>

‡ <http://hul.harvard.edu/jhove/>

§ <http://dioscuri.sourceforge.net/>

dokumentu vytvořeného ve starém textovém procesoru AmiPro s koncovkou .sam např. do PDF nebo do jiných v současnosti podporovaných formátů. Výhodou migrace je, že v případě, že uživatel chce uložený digitální objekt použít (zobrazit, spustit apod.), ten je v podstatě připraven k tomu být spuštěn v aktuálních verzích běžného softwaru na počítači uživatele. Nevýhodou je, že velmi často se migrací ztrácí část původního objektu (ne informace, ale typicky např. část formátování). Digitální objekt, který dostane budoucí uživatel, nemusí být zcela totožný s objektem původním.

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.

3.3 ZÁKLADNÍ FUNKCE DŮVĚRYHODNÉHO DIGITÁLNÍHO REPOZITÁŘE*

Máme-li hledat odpověď na otázku, proč je vybudování a provoz důvěryhodného digitálního repozitáře finančně i personálně tak náročné, že překračuje možnosti jednotlivých institucí, je třeba pochopit jeho základní funkce. V úplnosti je mapuje (a do detailů rozvíjí) referenční model OAIS. Digitální repozitář je definováno jako organizace, která uchovává informace s cílem jejich zpřístupnění a využití. 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 vypůjčit.

Digitální repozitář rovněž přijímá dokumenty (digitální) od producentů, zpravidla včetně dohodnutých metadat (soubor informací pro dodávání), tzv. SIP balíček (viz text výše). Výsledkem kontroly digitálních dokumentů a příslušných metadat a obohacení metadat (zde nehovoříme jen o metadatech popisných, ale i administrativních, technických a ochranných) je soubor informací pro archivaci – AIP balíček, při požadavku na zpřístupnění informací z repozitáře je generován soubor informací pro šíření – DIP balíček.

3.4 KRITÉRIA HODNOCENÍ DŮVĚRYHODNOSTI DIGITÁLNÍHO REPOZITÁŘE

Kritéria hodnocení důvěryhodnosti digitálního repozitáře byla původně obsažena ve dvou základních materiálech:

- *Trusted Digital Repositories : Attributes and Responsibilities* (RLG, OCLC, 2002)
- *Trustworthy Repositories Audit & Certification : Criteria and Checklist* (OCLC, CRL, 2007)

Starší z obou dokumentů vydaný v roce 2002 definuje obecné vlastnosti důvěryhodného repozitáře takto následovně > **důvěryhodný repozitář 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 repozitář, ale i pro svěřené digitální informace.
- Prokázat finanční zajištění v současnosti i trvale udržitelný rozvoj.
- Navrhnout systém pro správu digitálního repozitáře 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 repozitáři deponují.
- Disponovat strategií, pracovními postupy a službami, které umožňují snadné hodnocení a měření.

* STOKLASOVÁ, Bohdana; HUTAŘ, Jan. Nové směry v dlouhodobém uchovávání dokumentů... s. 87.

Navazující dokument *Trustworthy Repositories Audit & Certification : Criteria and Checklist* již představuje velmi detailní a propracovaný systém kriterií pro hodnocení důvěryhodného repozitáře, které se rozpadá do tří hlavních sekcí:

- a) Organizace (řízení, struktura, udržitelnost, finance)
- b) Správa digitálních objektů
- c) Technologie, technická infrastruktura, bezpečnost

V jednotlivých sekcích jsou sledovány okruhy, v jejichž rámci je třeba zodpovědět poměrně konkrétní otázky. Z uvedeného výčtu je zřejmé, že pro vybudování a provoz důvěryhodného repozitáře zdaleka nestačí zakoupení drahého technického a programového vybavení, právě naopak. 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 repozitáře. Musí být adekvátně finančně a personálně zajištěn nejen pro dobu vzniku repozitáře a po krátkou dobu po něm, ale dlouhodobě. Tvorba a provoz důvěryhodných digitálních repozitářů 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 a vědeckovýzkumných institucí v národním i mezinárodním kontextu*.

3.5 SOUČASNÉ NÁSTROJE NA CERTIFIKACI A AUDIT DŮVĚRYHODNÝCH DIGITÁLNÍCH REPOZITÁŘŮ

1) DRAMBORA (Digital Repository Audit Method Based on Risk Assessment[†])

- vytvořena ve spolupráci Digital Curation Centre (DCC) a DigitalPreservationEurope (DPE),
- společné body s TRAC a Nestor (viz níže),
- nemá za cíl udělit někomu certifikát, jde jen o nástroj pro samostatný audit, který si provede sama instituce,
- v případě zájmu pak lze provést následnou certifikaci zvenčí (např. TRAC).

2) Trustworthy Repositories Audit & Certification : Criteria and Checklist (TRAC)[‡]

- zdůrazňuje best practice kritéria pro důvěryhodné repozitáře ve třech různých oblastech,
- bere OAIS za své východisko a vztažný bod pro úspěšné porovnávání,
- aspiruje na to stát se standardem,
- klade důraz na proces certifikace,
- certifikace probíhá na objednávku, třetí nezávislou stranou (auditorem),
- proces je placený – zatím probíhá pouze v USA,
- ideální je navázat na self audit (DRAMBORA).

* STOKLASOVÁ, Bohdana; HUTAŘ, Jan. Nové směry v dlouhodobém uchování dokumentů... s. 87.

† <http://www.repositoryaudit.eu>

‡ <http://www.crl.edu/PDF/trac.pdf>

3) **NESTOR Criteria Catalogue***

- obsahuje 14 kritérií doplněných detailním vysvětlením a konkrétními příklady,
- dělí celou problematiku do skupin,
 - organizační rámec
 - object management
 - infrastruktura a zabezpečení
- odráží německý kontext (právní, finanční apod.).

4) **International Audit and Certification Birds of a Feather Group†**

- mezinárodní snaha o vytvoření ISO standardu, podle kterého by se v budoucnu prováděl kompletní audit a certifikace repozitáře,
- berou v potaz existující nástroje (viz nahoře) a snaží se je racionalizovat do jednoho dokumentu nebo souboru propojených dokumentů (vytvářejí různá srovnání, crosswalks apod.),
- jsou otevření všem zájemcům, každý se může zúčastnit!

Na základě těchto nástrojů lze definovat deset základních principů důvěryhodnosti, které jsou jakýmsi jejich průnikem. Mohou být prvotním východiskem, jakýmsi „samoodhadem“ předtím, než se opravdu přistoupí k serióznímu auditu nebo procesu certifikace‡. Těchto deset bodů můžeme považovat za desatero jakéhokoliv repozitáře, které chce být považováno za důvěryhodné (jaro 2007). Takový archiv:

1. se musí zavázat k neustálému opatrování/správě digitálních objektů pro určitou cílovou komunitu.
2. musí prokázat svou životaschopnost/způsobilost (včetně financování, personálních otázek, struktury, procesů), aby dostal stanovenému závazku.
3. musí si osvojit a dodržovat potřebná smluvní a zákonná práva a dostát všem z nich plynoucím závazkům.
4. musí mít efektivní a dostačující rámcovou strategii.
5. získává a ukládá digitální objekty na základě stanovených kritérií, které odpovídají cílům a schopnostem instituce.
6. neustále udržuje/zajišťuje integritu, autenticitu a využitelnost digitálních objektů, které trvale uchovává.
7. vytváří a uchovává potřebná metadata o událostech souvisejících s uloženými digitálními objekty v průběhu jejich uchování, jako i metadata o samotném vytvoření digitálních objektů, podmínkách zpřístupnění a kontextu využití digitálních objektů.
8. naplnit nezbytné požadavky na zpřístupnění objektů ven z repozitáře určité komunitě.
9. musí mít strategii pro plánování ochrany a souvisejících procesů včetně záchranných prací.
10. musí mít technickou infrastrukturu adekvátní pro účel neustálé údržby a zajištění digitálních objektů.

* <http://edoc.hu-berlin.de/series/nestor-materialien/8/PDF/8.pdf>

† www.digitalrepositoryauditandcertification.org

‡ Digital Curation Centre and Digital Preservation Europe. *DCC and DPE Repository Audit Method Based on Risk Assessment*, v. 1.0. Glasgow : DPE, 2007. Přístup z WWW: <<http://www.repositoryaudit.eu>>.

Pakliže je repozitář schopen:

- 1) rozpoznat a priorizovat hrozby, které ohrožují jejich aktivity;
- 2) vypořádat se s hrozbami tak, aby snížily možnost jejich výskytu;
- 3) stanovit nepředvídatelné události, aby se snížil efekt hrozby, kterou představují;

je velmi pravděpodobně připraven obdržet status důvěryhodnosti.

Důvěryhodnost je klíčovou vlastností, kterou musí certifikovaný repozitář demonstrovat. Digital preservation je v podstatě uvědomování si organizačních, procedurálních, technologických a jiných nejistot a jejich přeměna zvladatelná (řešitelná) rizika. *Pro získání takové důvěryhodnosti má právě napomoci nezávislý audit.*

3.5.1 Kontext existujících standardů

Všechny snahy o standardizaci procesu certifikace repozitářů berou v potaz již existující standardy:

- ISO 9000 (Quality assurance)
- ISO 17799 & 27001 (Information security)
- ISO 15489 (Institutional Records Management)
- ISO 14721 (Reference Model for an Open Archival Information system)
- COBIT 4.1

3.6 SELF AUDIT POMOCÍ NÁSTROJE DRAMBORA

Online nástroj DRAMBORA^{*} (Digital Repository Audit Method Based on Risk Assessment) vznikl ve spolupráci DCC[†] a projektu DPE[‡]. DRAMBORA nevznikla jako další certifikační nástroj, ale jako nástroj, který má pomoci instituci, která plánuje certifikaci svého repozitáře. Pomoci tím, že si instituce provede tzv. self audit. Může to udělat pracovník té instituce, musí ovšem být s DRAMBOROU seznámen. Tento „samoodhad“ může do velké míry snížit náklady na následnou externí certifikaci, která je placená (např. pro TRAC se hovoří o desítkách tisíc dolarů) a také, což je hlavní, ji může významně urychlit. DRAMBORA je od jara 2008 dostupná v podobě interaktivního nástroje, který je přístupný online a umožní projít všemi kroky auditu, ukládat hodnoty a vytvořit hodnocení[§].

DRAMBORA je myšlena jako nástroj, pomocí kterého může auditor odhadnout možnosti repozitáře, identifikovat jeho slabé stránky a rozpoznat jeho stránky silné. Toto je důležité zvláště v situaci, kdy se o repozitářích velmi mluví, spousta institucí si je zřizuje, ovšem zároveň jsou nynější repozitáře a jejich vývoj, měřeno z hlediska lidského věku, někde před pubertou. DRAMBORA je strukturována jako možná odpověď nebo pomůcka pro další vývoj repozitářů a hlavně pro jejich funkčnost v nedaleké i daleké budoucnosti. Soustřeďuje se na nich velká část našeho současného vědění, je tedy nezbytně nutné toto vědění zachovat.

3.6.1 Struktura celého procesu auditu

DRAMBORA má 6 fází:

1. identifikace východisek (status, kontext organizace/instituce apod.)
2. identifikace strategie a regulačního rámce (koncepce, legislativa apod.)

* <http://www.repositoryaudit.eu/>

† <http://www.dcc.ac.uk/>

‡ <http://www.digitalpreservationeurope.eu/>

§ <http://www.repositoryaudit.eu/>

3. identifikace aktivit a prostředků
4. identifikace hrozeb souvisejících s aktivitami a prostředky
5. vyhodnocení hrozeb
6. zvládnutí hrozeb

Výhoda nástroje DRAMBORA je i v tom, že se audit může po určité době opakovat. To je velmi dobré v tom případě, že instituce si udělá první audit svého repozitáře, odhalí jeho slabiny, ohodnotí hrozby, jejich možné následky a přijme opatření typu „zlepšit to a ono do dvou měsíců“ a poté, za ty dva měsíce, může nechat repozitář projít znovu celým procesem a výsledky obou auditů porovnat.

3.6.2 Výstup z celého DRAMBORA auditu a jeho výhody

Výstupem je mimo jiné zvýšená znalost o chodu repozitáře, struktuře procesů v něm, dále seznam hrozeb, které repozitáři hrozí. Záleží na auditorovi, jak podrobný tento seznam bude a také na instituci, jak kvalitní podklady a podporu auditorovi poskytne, což může ovlivnit jeho pochopení celého chodu repozitáře, potažmo instituce.

Provozovatel repozitáře získá ucelený pohled na všechny procesy, které mají s repozitářem souvislost a to ve formě bodových skóre. Ukáží se tak silné a naopak nejslabší oblasti, které mohou znamenat hrozbu. Někdy výsledky mohou být překvapivé, velmi často se za nejrizikovější oblast považují technologie, ovšem daleko větší riziko může být v nedostatečně zvládnutém a nejistém financování repozitáře nebo instituce, která jej provozuje nebo také např. v personálních otázkách (nedostatek odborníků díky nízkým platům apod.). To jsou věci, které mohou potenciálně ochromit chod celého systému stejně fatálně jako technická závada.

Výstupem ale bude také připravenost na kompletní externí audit (certifikaci). Důvodem proč by instituce mohla chtít dělat interní audit, může být i pouhé ověření, zda všechny procesy a celková strategie jsou nastaveny dobře a systém pracuje optimálně.

4 ZÁVĚR

Jak je patrné z textu, je problematika uchovávání a s tím spojené ochrany digitálních informací velmi aktuální a živá. Je to zapříčiněno mj. i tím, že instituce již čelí problémům vyplývajícím z používání digitálních informací (objektů). Paradoxní je, že většina uživatelů těchto informací v digitální podobě vidí pouze jejich výhody a nemá tušení o vznikajících hrozbách a komplikacích. Alarmující je, že se podobně chovají i někteří producenti i instituce, které digitální informace používají. To by se mělo díky nástrojům, koncepcím nebo i auditům velmi rychle změnit. Můžeme říci, že dnes se přesouvá těžiště od problematiky „jak vytvářet digitální informace“ ke stupni „jak digitální informace uchovávat v dlouhodobém časovém horizontu.“

Že jde o problematiku dneška, dokládá stoupající počet národních i mezinárodních aktivit, konferencí, odborných publikací apod. Může nás těšit, že Česká republika není tentokrát pozadu a je od počátku součástí těchto snah.

5 POUŽITÁ LITERATURA

1. American Library Association. *Definitions of Digital Preservation* [online]. Washington : American Library Association, c2008 , September 11, 2008 [cit. 2007-11-18]. Příklad z WWW: <<http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.cfm>>.
2. CARPENTER, Leona. *Repositories in Context: Digital Repositories as components of an integrated infrastructure for education* [online]. UKOLN, 2005 [cit. 2007-11-18]. Příklad z WWW: <<http://www.ukoln.ac.uk/events/delos-rep-workshop/presentations/carpenter.ppt>>.
3. *Catalogue of Criteria for Trusted Digital Repositories*. Version 1. Goettingen : Nestor Working Group, 2006. 48 s. Příklad z WWW: <<http://edoc.hu-berlin.de/series/nestor-materialien/8/PDF/8.pdf>>.
4. CCSDS 650.0-B-1: *Reference Model for an Open Archival Information System (OAIS)*. Blue Book. Issue 1. January 2002. 148 s. Příklad z WWW: <<http://nost.gsfc.nasa.gov/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>>.

5. Digital Curation Centre and DigitalPreservationEurope. DCC and DPE Repository Audit Method Based on Risk Assessment, v. 1.0. Glasgow : DPE, 2007. Přístup z WWW: <<http://www.repositoryaudit.eu>>.
6. GIARETTA, David. *ISO/CCSDS Open Archival Information System (OAIS) Reference Model* [online]. [s.l.] : [s.n], [s.a] [cit. 2007-11-18]. Přístup z WWW: <<http://www.sstd.rl.ac.uk/ccsdsp2/isoas/bnsc97/oais1/OAIS1.PPT>>.
7. KÖNIGOVÁ, Marie. *Vybrané kapitoly z informační vědy*. In *Informační studia a knihovnictví v elektronických textech* [elektronický zdroj]. I. Praha : UISK, 2002.
8. LAVOIE, Brian F. *The Open Archival Information System Reference Model: Introductory Guide*. DPC and OCLC Online, January 2004. 20 s. DPC Technology Watch Series Report 04-01. Přístup z WWW: <http://www.dpconline.org/docs/lavoie_OAIS.pdf>.
9. *The Orange Book. Management of Risk - Principles and Concepts*. Norwich : HM Treasury, 2004. 58 s. Přístupné z WWW: <<http://www.hm-treasury.gov.uk/media/FE6/60/FE66035B-BCDC-D4B3-11057A7707D2521F.pdf>>.
10. *Preservation Metadata and the OAIS Information Model. A Metadata Framework to Support the Preservation of Digital Objects*. A Report by The OCLC/RLG Working Group on Preservation Metadata. OCLC, 2002. 54 s. Přístup z WWW: <http://www.oclc.org/research/projects/pmwg/pm_framework.pdf>.
11. SAWYER, Don; REICH, Lou. *ISO "Reference Model For an Open Archival Information System (OAIS)"* [online]. Tutorial Presentation. Library of Congress, June 13, 2003. [cit. 2007-11-18]. Přístup z WWW: <http://nost.gsfc.nasa.gov/isoas/presentations/oais_tutorial_200306.ppt>.
12. STOKLASOVÁ, Bohdana. *Dlouhodobé uchování dokumentů v českých knihovnách* [online]. ISSS 2007, Hradec Králové 2.4. 2007 [cit. 2007-11-18]. Přístup z WWW: <<http://www.ndk.cz/publikace/iss-2007/nkcrstoklasova.ppt>>.
13. STOKLASOVÁ, Bohdana; HUTAŘ, Jan. *Nové směry v dlouhodobém uchování dokumentů v mezinárodním kontextu*. In *Automatizace knihovnických procesů 11*. Liberec 16.-17.5.2007. Praha : ČVUT, 2007. 136 s. ISBN 978-80-01-03691-4.
14. *Trustworthy Repositories Audit & Certification : Criteria and Checklist*. Dublin, OH : OCLC, CRL 2007. 94 s. Přístup z WWW: < <http://www.crl.edu/PDF/trac.pdf>>.

6 RELEVANTNÍ PROJEKTY A JEJICH WEBOVÉ STRÁNKY

- DigitalPreservationEurope

<http://www.digitalpreservationeurope.eu/registries/>

- CASPAR

Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval
<http://www.casparpreserves.eu/>

- Planets

Preservation and Long-term Access through Networked Services <http://planets-project.eu/>

- Nestor

<http://nestor.sub.uni-goettingen.de/>

- DCC

Digital Curation Centre <http://www.dcc.ac.uk/>

- **DPC**

Digital Preservation Coalition <http://www.dpconline.org/>

Slovník terminologie <http://www.dpconline.org/graphics/intro/definitions.html>

- **DRAMBORA**

Digital Repository Audit Method Based on Risk Assessment <http://www.repositoryaudit.eu/>

- **NDIIPP**

National Digital Information Infrastructure and Preservation Program
<http://www.digitalpreservation.gov/>

- **NDK**

Národní digitální knihovna <http://www.ndk.cz/>