Colin Rosenthal,
Asger Blekinge-Rasmussen,
Jan Hutař a kol.

Průvodce
plánem důvěryhodného digitálního repozitáře (PLATTER)

Národní knihovna
České republiky
National Library
of the Czech Republic

Průvodce plánem důvěryhodného digitálního repozitáře PLATTER

Národní knihovna
České republiky
National Library
of the Czech Republic

digital preservation Europe

Information Society
Technologies

Průvodce plánem důvěryhodného digitálního repozitáře

PLATTER

Colin Rosenthal,
Asger Blekinge-Rasmussen,
Jan Hutař a kol.

Průvodce plánem důvěryhodného digitálního repozitáře (PLATTER)

Praha 2009

Národní knihovna České republiky

KATALOGIZACE V KNIZE – NÁRODNÍ KNIHOVNA ČR

Rosenthal, Colin
Průvodce plánem důvěryhodného digitálního repozitáře (PLATTER) / Colin Rosenthal, Asger Blekinge-Rasmussen, Jan Hutař a kol. ; [překlad Jan Hutař, Ladislav Cubr, Marek Melichar]. – 1. vyd. – Praha : Národní knihovna ČR, 2009
Název originálu: Repository planning checklist and guidance
Přeloženo z angličtiny
ISBN 978-80-7050-569-4

930.25:004.056.3 * 930.25:004.08 * 004.6.056.5-022.314
- PLATTER
- digitální archivy
- dlouhodobá ochrana digitálních dat
- informační publikace

Publikace vznikla v rámci projektu DigitalPreservationEurope, který podpořila Evropská komise v 6. rámcovém programu, tematická priorita Information Society Technologies (IST-2005-2.5.10)

1. vydání

Colin Rosenthal, Asger Blekinge-Rasmussen, Jan Hutař, Andrew McHugh, Stephan Strodl, Emily Witham, Seamus Ross

Z anglického originálu Repository Planning Checklist and Guidance
přeložili: Jan Hutař, Ladislav Cubr, Marek Melichar
Odborná revize překladu: Andrea Fojtů

ISBN 978-80-7050-569-4

ÚVOD K ČESKÉMU VYDÁNÍ PLATTERU

Co je to „dlouhodobá ochrana digitálních dat“?

Objem digitálních dat, který je třeba dlouhodobě archivovat, se zhruba od poloviny devadesátých let minulého století začal prudce zvyšovat. Exponenciální nárůst digitálních objektů je patrný nejen v oblasti průmyslové, technické a telekomunikační. Archivy digitálních dat potřebují také oblasti, jako jsou medicína, státní správa, justice nebo policie. Digitalizace je dnes jednou ze základních strategií dlouhodobé ochrany analogových sbírek (tj. fyzických dokumentů a sbírkových předmětů) ve většině paměťových institucí. Důležitou oblastí pro ochranu digitálních dat jsou také univerzity a další výzkumné instituce. Jejich digitální repozitáře obsahují odborné publikace, kvalifikační práce studentů nebo digitální kopie fyzických dokumentů a také obrovské množství digitálních dat generovaných obory tzv. exaktních věd, často velmi specializovaných a využívajících složitých formátů. Všechna tato data musí být dlouhodobě dostupná a srozumitelná.

Problematika uchovávání digitálních dat v takové podobě, aby byla tato data znovu vyhledatelná, srozumitelná a použitelná i po dlouhé době a v jiných kontextech (na jiných platformách apod.), než v kterých původně vznikla, se anglicky obvykle nazývá „digital preservation“, „preservation of digital objects“ a někdy také „digital curation“.

Dlouhodobá ochrana digitálních dat je oblastí výzkumu a praxe. Není to zatím obor, který by bylo možné studovat na univerzitě, ani to není nějaký soubor softwarových nástrojů nebo procedur, které by bylo možné k ochraně dat použít.

Společným terminologickým a konceptuálním rámcem všech projektů v této oblasti je model OAIS *(Open Archival Information System)*, který původně vznikl v oblasti kosmického výzkumu. Dnes jsou základní pojmy tohoto modelu společným jazykem všech, kteří se ochranou digitálních dat zabývají.

Digitálními daty (informacemi, objekty) se zde myslí jednak produkty digitalizace původně analogových (fyzických) dokumentů, jednak objekty, které už jako digitální vznikly (tzv. *born digital*). Digitální data jsou uložena a spravována v digitálním repozitáři. Digitální repozitář (též digitální archiv) lze chápat jako organizaci lidí a systémů se závazkem ochraňovat a zpřístupňovat digitální data pro určitou skupinu uživatelů[1]. Infrastruktura jednotlivého repozitáře může být navržena tak, aby navíc umožňovala i snadnou vzájemnou spolupráci s jinými systémy (portály, digitálními knihovnami) pomocí daných komunikačních protokolů. Digitální knihovnou rozumíme soubor aplikací, které nad daty v repozitáři provádějí nějaké stanovené operace za účelem jejich zpřístupnění koncovým uživatelům. Digitální repozitáře sehrávají klíčovou roli v dlouhodobé ochraně digitálních dat.

V literatuře najdeme různé definice dlouhodobé digitální ochrany:

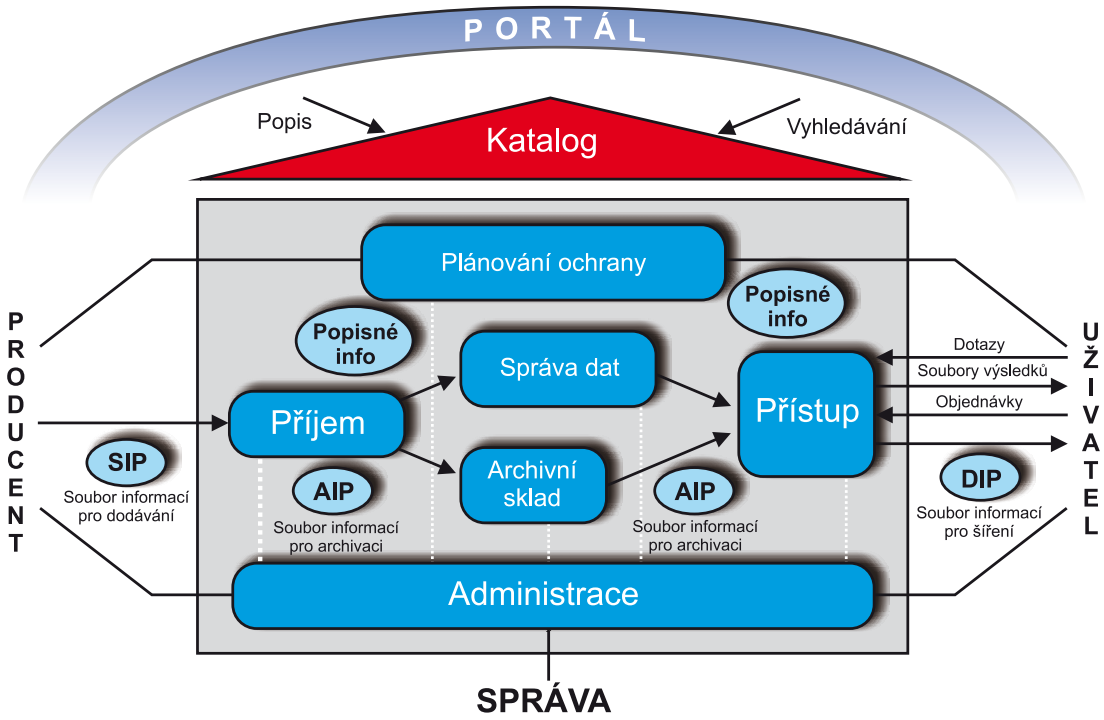
1. Řízená snaha zajistit použitelnost digitálních objektů po mnoho let. Jde především o snahu zajistit, aby se digitální objekty nikdy neztratily, aby nedošlo k jejich poškození, aby bylo možné je vždy najít a aby byly srozumitelné. To vše bez ohledu na jejich možnou technologickou zastaralost[2].
2. Uchovávání informací v přesné a samostatně srozumitelné podobě po dlouhou dobu[3].
3. Všechny aktivity zabývající se správou digitálních nebo elektronických objektů spojené s jejich uchováváním a zpřístupňováním[4].

Dlouhodobou ochranou se obvykle myslí ochrana během (někdy i za) období, ve kterém dochází k technologickým změnám (jako jsou zavádění nových formátů, nástupy nových médií nebo nové architektury procesorů, počítačů apod.).

Jak již bylo zmíněno, digitální objekty jsou v rámci této ochrany uloženy v nějakém digitálním repozitáři, který tuto ochranu zajišťuje. Zároveň jsou však tyto objekty dále využívány určitou komunitou uživatelů, a tak musí být zajištěna nejen ochrana objektů, ale také ochrana přístupu k těmto objektům při dodržení zákonných omezení vyplývajících z autorských práv a dalších možných omezení jejich použití.

1 CCSDS. *Reference Model...* 2002, s. 1–1.
2 GLADNEY, H. M. *Preserving digital information...* 2007, s. 270.
3 CCSDS. *Reference Model...* 2002, s. 1–11.
4 VERHEUL, I. *Networking for Digital Preservation...* 2006, s. 20.

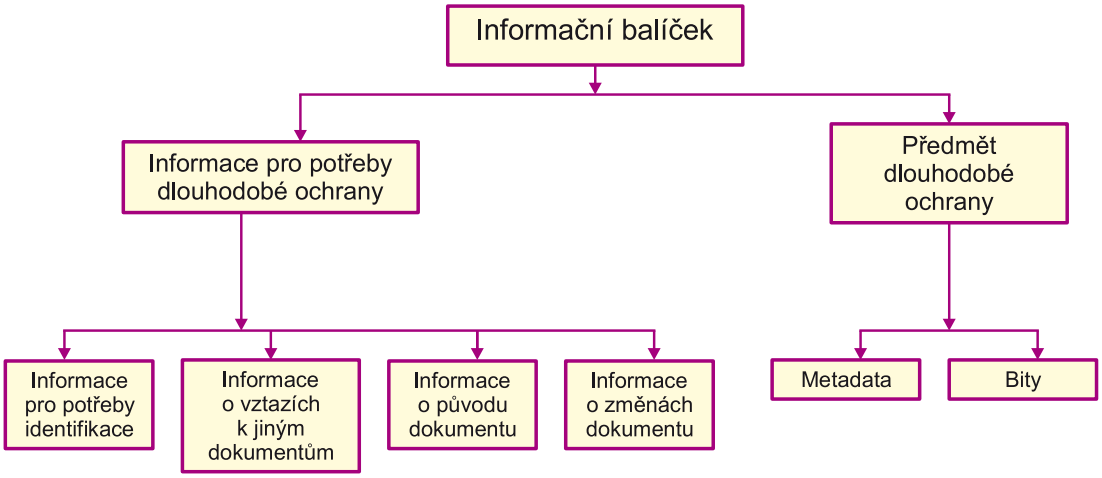
Referenční model OAIS (Open Archival Information System) – ISO 14721



Dlouhodobá ochrana digitálních dat je proces, jehož cíle mohou být odlišné pro různé typy digitálních objektů nebo v různých kontextech. Nejnižší úrovní dlouhodobé ochrany digitálních dat je snaha udržet digitální objekty v podobě původních a nijak nezměněných bitových sekvencí bez ohledu na srozumitelnost nebo na použitelnost těchto bitů. Komplexnější přístup k dlouhodobé ochraně digitálních dat se snaží zajistit, aby byly informace v budoucnosti dostupné v takové podobě, která bude čitelná a srozumitelná budoucím uživatelům. Tj. aby význam, který digitální objekty nesou (v modelu OAIS *content information*), zůstal zachován a neztratil srozumitelnost. Tento přístup již vyžaduje použití technických, strukturálních i jiných metadat a jejich ochranu a v řadě případů také uchovávání dalších informací, které budoucím uživatelům umožní porozumět obsahu digitálních objektů (v modelu OAIS se všechna tato metadata a další informace, které pomáhají poznat a pochopit původní strukturu, obsah a smysl dokumentů, nazývají *representation information*). Pro uživatele může být také důležité mít k dispozici informace o původu každého uchovávaného objektu (v modelu OAIS *provenance information*) a o operacích, které archiv nebo repozitář s daty během doby jejich uložení provedl (v modelu OAIS *fixity information*). Uživatelé musí mít důvěru v to, že objekty, tak jak je z repozitáře dostávají, jsou autentické, přesné, úplné, tj. právě v takovém stavu, v jakém byly do repozitáře vloženy, resp. v takovém stavu, v jakém je stanoveno (např. v *representation information*), že mají být. Říkáme, že digitální repozitář, který je schopen dostát všem těmto nárokům, je tzv. důvěryhodným digitálním repozitářem. Důvěryhodnost je tedy přidanou hodnotou, kterou repozitář získá, pokud je navržen a provozován podle určitých kritérií.

Z technického hlediska existuje řada přístupů nebo metod, které dlouhodobou ochranu digitálních dat usnadňují (zálohování, vícenásobné kopie, pravidelné výměny úložných médií, analogové zálohy, emulace, migrace, udržování dokumentace, audit repozitáře, sledování rizik, hodnocení formátů, sledování softwaru a hardwaru, pečlivá tvorba metadat pro potřeby dlouhodobé ochrany atd.). Ovšem dlouhodobá ochrana digitálních dat je technickým problémem jen zčásti. Dalším problémem je, jak zajistit organizaci, financování, kvalifikovaný personál a efektivní řízení.

Informační balíček podle modelu OAIS



Výzkumné projekty v této oblasti

Ve světě probíhala a probíhá řada výzkumných projektů, jejichž cílem je posunout znalosti v oblasti dlouhodobé ochrany digitálních dat o krok dál. Mimo Evropu jsou v této oblasti obzvláště aktivní na Novém Zélandě a v Austrálii a pochopitelně ve Spojených státech amerických. V Evropě je řada projektů financována prostřednictvím rámcových programů EU. V šestém rámcovém programu (FP6) byly financovány projekty týkající se dlouhodobé ochrany digitálních dat v rámci programu DigiCult – především bychom měli zmínit projekty jako DELOS, PLANETS, CASPAR a DPE[5]. V rámci posledního zmíněného projektu vznikal zde přeložený dokument, PLATTER. V sedmém rámcovém programu jsou financovány projekty jako SHAMAN, LIWA a další[6]. Kromě toho existují i národní projekty v oblasti dlouhodobé ochrany digitálních dat (v Německu především projekt Nestor, projekty Nizozemské královské knihovny nebo britské projekty financované z programu JISC). Vesměs se jedná o projekty výzkumné, ovšem mnohé z nich tvoří i praktické nástroje, které mohou usnadnit práci lidem starajícím se o chod a správu digitálního repozitáře (PLANETS testbed, PLANETS PLATO, Caspar repinfo tools).

Potřeby dlouhodobé ochrany digitálních dat jsou v poslední době také zohledňovány ve speciálních softwarových nástrojích určených pro správu a provoz digitálních knihoven a repozitářů, ať už jde o komerční software, nebo nástroje vytvořené jako open source.

Důležitou součástí dlouhodobé ochrany digitálních dat je certifikace repozitářů, jíž by měl předcházet (řádný) audit. V této oblasti v současnosti existují tři hlavní projekty/nástroje – Nestor catalogue, TRAC a DRAMBORA.

Jak PLATTER vznikl a komu je určen?

Plán důvěryhodného digitálního repozitáře (PLATTER), který zde předkládáme českým čtenářům, vznikl v rámci projektu DigitalPreservationEurope (DPE). Projekt DPE má za cíl popularizovat problematiku dlouhodobé ochrany digitálních dat mezi širší odbornou veřejností a koordinovat spolupráci jednotlivých výzkumných snah v této oblasti. Kromě PLATTERu se DPE podílí na vývoji DRAMBORA – nástroje pro sebehodnocení repozitářů. Internetová stránka DPE obsahuje mj. seznamy digitálních repozitářů, prezentací, výzkumných projektů a zpráv.

Cílem PLATTERu je poskytnout čtenáři základní orientaci v problematice dlouhodobé ochrany digitálních dat v repozitáři. PLATTER není technickou příručkou pro programátory, kteří by chtěli budovat důvěryhodný digitální repozitář. Je spíš kuchařkou pro manažery či pracovníky zodpovědné za digitální data. Ukazuje rozsah problematiky a upozorňuje na možná rizika a problémy. PLATTER je zcela prakticky

5 http://cordis.europa.eu/fp7/ict/telearn-digicult/digicult-projects-fp6_en.html
6 http://cordis.europa.eu/fp7/ict/telearn-digicult/digicult-call1_en.html

orientovaný dokument. Snaží se respektovat současný stav znalostí v oblasti dlouhodobé ochrany digitálních dat, ale zároveň by měl být srozumitelný i laikovi.

Nejvýhodnější je používat PLATTER při vytváření projektu digitálního repozitáře. Je dobré si v každém bodě četby PLATTERu uvědomit specifika vlastní situace, najít relevantní cíle nebo prostředky k jejich dosažení, specifikovat zdroje financování i zdroje možných rizik. Cílem PLATTERu je být prvním dokumentem, který vezmeme do ruky před tím, než začneme budovat repozitář nebo než se začneme snažit o jeho důvěryhodnost.

PLATTER je určen následujícím typům uživatelů, čtenářů:

- manažerům a vedoucím pracovníkům institucí, které spravují sbírky digitálních objektů,
- zaměstnancům IT oddělení těchto institucí,
- studentům informačních studií nebo informačních technologií, kteří mají zájem o problematiku archivace dat,
- všem, kdo se zabývají digitalizací jakéhokoli typu analogového materiálu,
- laikům, kteří mají doma ve své soukromé sbírce velké množství digitálních objektů, jako jsou třeba fotografie, multimédia nebo elektronické texty.

Další zdroje informací:

Digital Preservation Management Resources
http://www.icpsr.umich.edu/dpm/index.html

TRAC
http://www.crl.edu/PDF/trac.pdf

Nestor
http://edoc.hu-berlin.de/series/nestor-materialien/8/PDF/8.pdf

DRAMBORA
http://www.repositoryaudit.eu/

PLATTER
http://www.digitalpreservationeurope.eu/platter/

WePreserve
http://www.wepreserve.eu

Webová stránka DPE
http://www.digitalpreservationeurope.eu/registries/resources
http://www.digitalpreservationeurope.eu/publications

DP Coalition
http://www.dpconline.org/

ERPANET
http://www.erpanet.org/

NDIIPP / LOC
http://www.digitalpreservation.gov/

PLANETS
http://www.planets-project.eu/

CASPARS
http://www.casparpreserves.eu

PADI
http://www.nla.gov.au/padi

FP 7
http://cordis.europa.eu/fp7/ict/telearn-digicult/digicult-preservation_en.html

SHAMAN
http://www.shaman-ip.eu

Experti, výzkumná centra

Webová stránka DPE – registry
http://www.digitalpreservationeurope.eu/registries/

Výzkumná centra

DP Coalition
http://www.dpconline.org

DCC
http://www.dcc.ac.uk/

Konference

iPRES
http://www.bl.uk/ipres2008

Konference WePreserve
http://www.wepreserve.eu/events/nice-2008/

Školicí střediska

DELOS
http://www.delos.info/

DPE
http://www.digitalpreservationeurope.eu/registries/trainers/

1 SHRNUTÍ A ÚVOD DO PLATTERU

Cílem tohoto dokumentu je představit nástroj pro plánování důvěryhodného digitálního repozitáře (*Planning Tool for Trusted Electronic Repositories* – dále jen PLATTER). PLATTER pomáhá plánovat cíle, úkoly a předpokládanou výkonnost digitálního repozitáře v různých fázích jeho budování tak, aby mohl být všemi stakeholdery[7] považován za důvěryhodný.

PLATTER není nástrojem auditu nebo certifikace, je spíše jejich doplňkem. Má pomoci provozovatelům vznikajících repozitářů zahrnout důvěryhodnost do svých cílů hned od počátku plánování. Repozitář, který bude naplánován s použitím PLATTERu, bude v lepší pozici při auditu jeho schopnosti dlouhodobě zajišťovat použitelnost a dostupnost dat.

Vzhledem k omezenému rozsahu tohoto dokumentu se soustředíme pouze na proces definice cílů repozitáře. Management realizace těchto cílů, zahrnující tak odlišné oblasti jako jsou financování, řízení lidských zdrojů, plánování softwaru a hardwaru, vytváření datových skladů atd., je příliš velkým soustem pro jediný dokument a vyžaduje pomoc řady různých specialistů.

Při vytváření všeobecného nástroje pro formulaci cílů a úkolů digitálního repozitáře se musíme vypořádat se značnou různorodostí organizací, kterým lze říkat „digitální repozitář“. PLATTER tuto variabilitu zohledňuje: každý repozitář nejprve odpovídá na klasifikační dotazník, který ho charakterizuje v porovnání s jinými repozitáři, a pomáhá zjistit, zda jsou vytyčené cíle v dané situaci uskutečnitelné.

Proces PLATTERu je založen na souboru plánů strategických cílů *(Strategic Objective Plans)*. V nich repozitář stanoví úkoly, cíle nebo klíčové indikátory v oblastech rozhodujících pro dosažení důvěryhodnosti. PLATTER bude základem elektronického nástroje, pomocí něhož budou repozitáře moci porovnat svoje cíle s cíli jiných podobných repozitářů. Plány strategických cílů by měly být živými organizmy, které se vyvíjejí spolu s repozitářem. PLATTER určuje plánovací cyklus, ve kterém se plány strategických cílů budou měnit společně s organizací provozující repozitář.

7 Stakeholdery zde rozumíme institucionální nebo individuální aktéry, kterých se budování, provoz a činnost digitálního repozitáře jakkoli dotýká (například zřizovatele, vlastníky dat, držitele autorských práv, uživatele, zaměstnance repozitáře atd.). Definice stakeholdera z oblasti projektového řízení zní: „jednotlivec nebo organizace, kteří se projektu aktivně účastní, nebo jejichž zájmy mohou pozitivně nebo negativně ovlivnit výsledky projektu a jeho dokončení. Stakeholdeři mají vliv na projekt a jeho výsledky.“ (A *guide to the project management body of knowledge.* Newtown Township (USA): Project Management Institute, 2000, str. 376.)

2 DŮVĚRYHODNÝ REPOZITÁŘ

Pojem „digitální repozitář“ se používá pro několik různých vzájemně se překrývajících systémů a organizací. Někdy označuje pouze soubory digitálních dat, které uplatňují konkrétní model, např. OAIS, nebo protokol, např. OAI-PMH[8]. V jiných kontextech se pojem repozitář používá velmi volně pro jakoukoli organizaci, která spravuje digitální obsah a zpřístupňuje ho určité skupině konečných uživatelů. Pojem digitální repozitář se také užívá, poněkud abstraktně, pro soubory služeb, které se zabývají získáváním, správou a zpřístupňováním digitálního obsahu. Tento poslední význam je užitečný zvláště v případech, kdy jsou služby poskytovány několika institucemi spolupracujícími v konsorciu.

Pojem „důvěryhodnost“ (anglicky *trust*, používá se také v gramaticky nepřesných tvarech *trustedness* nebo *trustworthiness*) je definován poněkud úžeji a přesněji. Repozitář je považován za důvěryhodný, pokud lze prokázat jeho schopnost plnit určité funkce a pokud tyto funkce splňují minimální dohodnutá kritéria, jež by měly splňovat všechny „důvěryhodné repozitáře“. Rozhodující je, že naplňování těchto kritérií musí být prokazatelné, což znamená, že dosažení důvěryhodnosti je do velké míry závislé na auditu a certifikaci.

Za tímto účelem byly vyvinuty v několika projektech nástroje pro audit a sebehodnocení repozitářů. Využívají v podstatě dvou komplementárních přístupů. TRAC[9] a Nestor[10] vytvořily dotazníky ověřující naplnění stanovených kritérií, jež musí repozitáře prokazatelně splňovat, aby mohly získat certifikát. Naproti tomu DRAMBORA[11] je návodem k sebehodnocení, který repozitářům umožňuje posoudit jejich schopnost dosahovat cílů, které si samy stanovily. Každá metoda má své slabé a silné stránky. První přístup, dotazníky ověřující naplnění kritérií, je konkrétnější, a proto je vhodnější pro certifikaci. Na druhé straně je poněkud rigidní a může být obtížné ho adekvátně použít ve všech různých typech digitálních repozitářů, které se chtějí stát důvěryhodnými. Naopak DRAMBORA je velmi pružný nástroj, protože porovnává výkon repozitáře s jeho vlastními cíli, a ne s nějakými zvnějšku definovanými standardy. To však znamená, že repozitář hodnocený pomocí DRAMBORy bude vždy jen tak dobrý, nakolik dostojí cílům, které si sám stanovil.

Kompromisem by bylo nechat repozitáře vybrat své vlastní cíle ze souboru základních požadavků, které jsou relevantní pro jakýkoli důvěryhodný repozitář. Takový soubor všeobecně uznávaných požadavků na důvěryhodný repozitář představuje například *Ten Core Principles of Trust Repository Design* (Deset základních principů důvěryhodného repozitáře), který vytvořily *The Center for Research Libraries* (CRL), *The Digital Curation Centre* (DCC), *DigitalPreservationEurope* (DPE) a *The German Network of Expertise in Digital long-term preservation* (Nestor) na setkání CRL v Chicagu v lednu 2007. Základní principy říkají, že repozitář:

1. se věnuje trvalé správě digitálních objektů pro definovanou komunitu / definované komunity
2. musí prokázat organizační způsobilost pro tento úkol (tzn. vhodné financování, personální zajištění a řízení)
3. dostojí smluvním a právním požadavkům a splní povinnosti z nich vyplývající
4. má vypracovanou účelnou a účinnou metodiku
5. získává a zpracovává digitální objekty podle stanovených kritérií, která odpovídají jeho cílům a schopnostem
6. udržuje a zajišťuje dlouhodobou integritu, autenticitu a použitelnost spravovaných digitálních objektů
7. archivuje potřebná metadata o všech akcích, které byly s digitálními objekty v průběhu jejich uložení provedeny; také shromažďuje související informace o vzniku, podpoře dostupnosti a využívání objektů před jejich vstupem do repozitáře
8. naplňuje potřebná kritéria pro zpřístupňování
9. má strategický program pro plánování ochrany

8 Open Archives Initiative. *The Open Archives Initiative Protocol for Metadata Harvesting* [online].
9 CRL; OCLC. *Trustworthy Repositories Audit & Certification...*2007.
10 Nestor. *Catalogue of Criteria...2006.*
11 *DRAMBORA interactive...* [online].

10. má odpovídající technickou infrastrukturu, potřebnou k trvalému udržování a zabezpečení spravovaných digitálních objektů[12].

Pro sebehodnotící nástroj DRAMBORA jsou tyto principy základem k identifikaci možných rizik. Pro nástroje certifikace představují dohodnuté klasifikační schéma těch položek, které mají být sledovány. Zbývá zajistit, a to je předmětem tohoto dokumentu, aby tyto principy byly brány v úvahu už v době plánování repozitáře, tak aby byl repozitář „připravený na důvěryhodnost“ *(trust-ready)* hned od počátku plánování.

12 CRL, et. al. *Core Requirements for Digital Archives*.

3 KLASIFIKACE REPOZITÁŘŮ

Jedním z hlavních úskalí na cestě k větší důvěryhodnosti je mimořádná různorodost organizací, pro které se označení repozitář používá. Praktické zkušenosti s použitím nástrojů DRAMBORA nebo TRAC ukazují, že důvěryhodnosti by chtělo dosáhnout mnoho různých typů repozitářů – například repozitáře národních knihoven a archivů, institucionální repozitáře, oborové repozitáře, archivy vědeckých dat – všude si uvědomují význam auditu jako nástroje identifikace silných a slabých stránek a jako nástroje externího hodnocení své práce.

Žádný univerzální přístup nemůže fungovat ve všech typech repozitářů. Je proto zásadní, aby ten, kdo repozitář plánuje, dokázal určit jeho typ a měl možnost srovnat svoje plány a praxi s jinými podobnými repozitáři. První fází PLATTERu je tedy klasifikace, která umožňuje repozitářům srovnání s repozitáři podobného typu. Lze to udělat několika způsoby, v PLATTERu například pomocí řady nezávislých os seskupených do čtyř hlavních tříd:

- účel a funkce repozitáře
- velikost repozitáře
- provoz
- technická řešení a možnosti implementace

Klasifikační osy jsou definovány velmi obecně tak, aby byly použitelné pro všechny typy repozitářů. Je ovšem pravděpodobné, že některé repozitáře shledají některé klasifikační osy pro popis svého fungování jako velmi omezující nebo zbytečné. Přesto očekáváme, že zde představená taxonomie, uchopena jako celek, bude platit pro ty repozitáře, které by dnes nebo v blízké budoucnosti mohly usilovat o důvěryhodnost.

3.1 Účel a funkce repozitáře

Cílem této části je určit typ repozitáře z funkčního hlediska. Požadavky na repozitář národní knihovny nebo národního archivu mohou být úplně jiné než požadavky kladené na institucionální nebo oborový repozitář či na repozitář vědeckých dat.

Otázka 1.1 Mandát, pověření

Repozitáře mají mandát z různých zdrojů. Národní knihovny a archivy mají obvykle mandát od vlády nebo od relevantního ministerstva. Mnoho repozitářů je součástí mateřské instituce, mají mandát od této instituce a zároveň se podílejí na mandátu mateřské instituce. Mnohé repozitáře, zvláště oborové nebo menší, definují svůj mandát samostatně.

Co je zdrojem mandátu repozitáře?	Např. vláda, mateřská instituce, sebe-definice.

Otázka 1.2 Status

Fungování repozitáře je do velké míry ovlivněno tím, zda se jedná o komerční projekt, který má přinášet svojí mateřské instituci nějaký zisk, nebo zda je neziskový.

Buduje se repozitář za účelem zisku, nebo má být neziskový?	

Otázka 1.3 Právní podmínky získávání obsahu

Při plánování repozitáře je také zásadní otázka právních podmínek získávání obsahu. Situace může být velmi odlišná v různých repozitářích. Záleží na tom, zda má repozitář všechna práva obsah získávat, nebo zda musí poskytnutí těchto práv dojednávat s jejich držiteli. Národní archivy obvykle alespoň část obsahu získávají na základě zákona o archivnictví, národní knihovny získávají materiál na základě právních předpisů o povinném výtisku atp. Ostatní typy repozitářů musí obvykle získávání materiálu zajistit smluvně nebo na základě dobrovolnosti přímo s poskytovateli obsahu.

Získává repozitář podstatnou část svého obsahu ze zdrojů, na které má ze zákona nárok (tj. z archiválií nebo povinných výtisků)?	

Otázka 1.4 Provozní vyzrálost

Plánuje-li se zcela nový repozitář nebo má-li projít již plně funkční repozitář změnou, proces plánování repozitáře se pro oba případy velmi liší. V případě nového repozitáře lze od začátku nastavit procesy a pravidla relativně libovolně s ohledem na důvěryhodnost. Navrhnout všechny procesy repozitáře tak, aby fungoval efektivně a zároveň při tom směřoval k druhému cíli, kterým je důvěryhodnost, je však velmi obtížné. Systém již fungujícího repozitáře může být efektivní a získání důvěryhodnosti bude možná vyžadovat jen relativně malé organizační změny. I ty však mohou jít proti organizační setrvačnosti, navíc je tu další nebezpečí, že se při tom objeví skryté nedostatky v procedurách repozitáře, jejichž náprava bude vyžadovat další podstatné a nákladné organizační změny.

Provozní vyzrálost repozitáře často nelze posoudit jako jediný parametr. Nejlepší je posuzovat provozní vyzrálost jednotlivých služeb repozitáře samostatně. PLATTER je spíše vhodný pro nově vznikající repozitáře. Ty repozitáře, které již nějakou dobu fungují a chtěly by se posunout směrem k důvěryhodnosti, obvykle mohou lépe využít analýzu rizik nástroje DRAMBORA.

Je již repozitář v provozu? (ještě ne; již ano, ale stále ho budujeme; již v plném provozu)	

3.2 Velikost repozitáře

Nyní se budeme věnovat různým faktorům, které ovlivňují celkový rozsah repozitáře, ať již z hlediska lidských zdrojů, technologií nebo financí. Deset základních principů důvěryhodného repozitáře (viz kapitola 2) ukazuje, že jak velké, tak malé repozitáře stojí před řadou stejných problémů, jejichž řešení jsou často závislá na jejich velikosti. Srovnání organizační a technické struktury jednotlivých repozitářů je přirozeně snazší mezi repozitáři podobné velikosti.

Otázka 2.1 a 2.2 Množství dat

Nejjednoduššími indikátory velikosti repozitáře jsou předpokládaný objem uložených dat a odhadovaný počet souborů nebo jednotlivých digitálních objektů. Celkový objem dat rozhodujícím způsobem ovlivňuje architekturu IT infrastruktury. Vliv počtu digitálních objektů nebo souborů nemusí být patrný na první pohled. V praxi však může mít počet digitálních objektů mimořádný vliv na náklady managementu dat. Složitost nakládání s daty roste s počtem spravovaných digitálních objektů, a nikoli s celkovou velikostí repozitáře. Množství metadat je například závislé na počtu jednotlivých digitálních objektů a ovlivňuje zásadním způsobem i používaný vyhledávací systém. Pokud je počet uložených objektů velmi vysoký, má to vliv na architekturu datového skladu; je například třeba soubory agregovat do formátů jako je ARC[13]. Repozitáře by se měly pokusit odhadnout tempo svého růstu v nejbližší budoucnosti. Jelikož jde pouze o odhad, má zároveň smysl specifikovat možnou chybu tohoto odhadu.

Jaký je předpokládaný objem digitálního materiálu, který budete ročně archivovat?	
Přibližně kolik samostatných digitálních objektů budete ročně archivovat?	

Otázka 2.3 a 2.4 Lidské zdroje

Dalším důležitým indikátorem velikosti repozitáře je počet zaměstnanců, kteří v něm pracují, a počet uživatelů jeho služeb. Obojí může být těžké odhadnout. Za zaměstnance považujeme osoby, které se práci v repozitáři věnují na plný úvazek. Tím vyřazujeme všechny možné zaměstnance firem najatých externě na různé práce v repozitáři. Srovnání mezi repozitáři, které využívají hodně outsourcingu,

13 BURNER, Mike; KAHLE, Brewster. *Arc File Format [online].*

a těmi, které preferují vlastní *(in-house)* řešení, je pak obtížnější. Přesto doporučujeme počítat pouze interní zaměstnance, protože a) odhadnout počet lidí pracujících v outsourcovaných oblastech může být složité, b) repozitáře významně využívající outsourcing a ty preferující vlastní řešení budeme (viz kap. 3.4) považovat za dostatečně odlišné a nebudeme je tedy navzájem přímo srovnávat.

Rovněž tak může být obtížné, zvláště pro nové repozitáře, změřit počet uživatelů. Pokud přijmeme premisu, že repozitář spravuje data pro nějakou konkrétní skupinu uživatelů, musí být nějaká forma průzkumu trhu zásadní součástí plánování důvěryhodného repozitáře. Měli bychom se pokusit odhadnout velikost komunity budoucích uživatelů repozitáře, i kdyby tento odhad měl být relativně nepřesný.

Kolik přesně zaměstnanců pracujících na plný úvazek repozitář zaměstnává?	
Kolik předpokládáte unikátních uživatelů vašeho repozitáře během jednoho kalendářního roku?	

3.3 Provoz

V následující skupině otázek se zabýváme tím, jak materiál do repozitáře vstupuje, jaký materiál repozitář uchovává a do jaké míry je tento materiál dostupný uživatelům. Parametry posuzované následujícími otázkami se mohou v jednotlivých repozitářích velmi odlišovat a neexistuje žádný apriorní důvod se domnívat, že spolu budou nějak korelovat. Zařadili jsme je do kapitoly „provoz“ proto, že spolu konceptuálně souvisí.

Otázka 3.1 Metody akvizice

Nejprve je potřebné si ujasnit, jak data do repozitáře vstupují. Data z externích zdrojů vstupují do repozitáře v zásadě dvěma způsoby: buď repozitář sám data aktivně získává, nebo jsou do něj zvnějšku vkládána. Tyto dva způsoby akvizice můžeme pojmenovat sklízená data *(pull)* a vkládaná data *(push)*. Třetím způsobem, jak může repozitář získávat data, je generování vlastních dat interní digitalizací. Pochopitelně, způsob akvizice dat má dalekosáhlé dopady na průběh celého procesu vkládání dat do repozitáře a na všechny související funkce jako validace dat, generování metadat a zajišťování kvality dat.

Kterým ze tří způsobů (push, pull, vlastní vytváření dat digitalizací) získává repozitář nejvýznamnější část svých dat?	

Otázka 3.2 Komplexnost dat

Digitální materiál existuje v mnoha formách. Problémy s dlouhodobým uchováváním a ochranou jsou zjevně větší u některých typů dat než u jiných. Charakterizovat komplexnost dat není snadné a může to odporovat našemu očekávání. Například formáty videa jsou většinou dobře popsané s dostatečnými specifikacemi, které jsou tak jednoduché, že umožňují video přímo přehrát. Naproti tomu některé textové formáty, jako například MS Word, jsou obalové formáty, které mohou obsahovat komplexní informace, jako jsou například tabulky nebo vnořené databáze. Repozitáři proto nestačí pouze definovat formát dokumentu, který přijímá, musí také vědět, jak moc komplexní data může očekávat nebo povolit ve svém archivu.

Pro zjednodušení jsme rozdělili dokumenty do tří skupin podle komplexnosti:
jednoduchá data: např. jednoduché textové formáty, obrázky nebo video
středně komplexní data: např. složené dokumenty s množstvím spojení
velmi komplexní data: např. software, texty s vloženými tabulkami s vnitřními vztahy

Je většina dat v repozitáři velmi komplexní, středně komplexní nebo jednoduchá?	

Otázka 3.3 Specializace dat

Specializací dat myslíme to, do jaké míry je k použití a interpretaci materiálu v repozitáři nutné mít expertní znalosti. Správné zhodnocení míry specializace digitálních dat je rozhodujícím krokem k úspěšnému rozhodování o strategii jejich ochrany. Specializace není komplexnost. Například rodinné fotografie a obrazy získané medicínskými zobrazovacími metodami mohou být daty stejně komplexními, pokud vyžadují stejný software a hardware. Ovšem specializace medicínských obrazů je mnohem vyšší, protože je potřeba velmi odborných znalostí například k rozhodnutí, jaká metadata mají být spolu s dokumentem uložena, jaké vlastnosti daného obrázku jsou nejdůležitější a musí být za každou cenu dlouhodobě udrženy.

Jak moc specializovaná jsou data v repozitáři? (velmi specializovaná, středně, málo)	

Otázka 3.4 Citlivost dat

Citlivostí dat myslíme soubor etických a právních otázek souvisejících se získáváním, ukládáním a zveřejňováním materiálu z repozitáře. Příkladem velmi citlivého materiálu jsou kupříkladu data, která mají velkou komerční hodnotu, nebo lékařská data obsahující osobní informace o konkrétních jednotlivcích. O tom, jakou úroveň zabezpečení musíme nastavit, rozhodují vždy nejcitlivější data v repozitáři.

Jak citlivý je nejcitlivější materiál v repozitáři (velmi citlivý, středně, málo)?	

Otázka 3.5 Oprávnění k přístupu

Materiál v repozitáři může být volně dostupný všem, volně dostupný omezené skupině uživatelů (například pouze badatelům), nebo zcela nedostupný. Mnoho repozitářů má data ve více než jedné z těchto kategorií.

Jaká data v repozitáři převládají (volně dostupná, omezeně dostupná nebo zcela nedostupná)?	

3.4 Technická řešení a možnosti implementace

Tato skupina otázek se zabývá možnými variantami implementace systému repozitáře.

Otázka 4.1 Zdroj metadat

Repozitář potřebuje ke své činnosti adekvátní bibliografická, popisná aj. metadata. Ta lze získávat z několika zdrojů:

- mohou být vkládána odděleně vkladatelem dokumentu, např. využitím HTTP hlavičky webové stránky
- mohou být manuálně získávána z dodaných dat
- mohou být automaticky extrahována z dodaných dat
- mohou být získána od třetí strany (dodavatel dat/depozitor)

Jaké jsou hlavní zdroje bibliografických a deskriptivních metadat v repozitáři?	

Otázka 4.2 Standardy interoperability

Zásadním vývojem v technologii repozitářů se staly standardy interoperability, které jsou pro provoz repozitáře klíčové. V technické rovině umožňují jednotlivým repozitářům nabízet (zviditelňovat) své zdroje sdílením služeb, datového materiálu a metadat. Identifikace formátů, nástroje validace, vyhledávání a zpřístupňování dat, automatická replikace za účelem ochrany, to jsou příklady služeb, kde

je interoperabilita důležitá. Jen málo standardů interoperability je dosud všeobecně přijímáno, což se ovšem v budoucnosti jistě změní.

Jaké standardy interoperability jsou v repozitáři zavedeny?	

Otázka 4.3 Strategie ukládání

Otázka se netýká technických podrobností architektury repozitáře, ale základní strategie ukládání dat, především toho, zda repozitář provozuje svůj vlastní datový sklad, nebo zda si datový prostor pronajímá od externího poskytovatele. Třetí variantou je ukládání ve vlastním datovém prostoru ovšem s externí podporou a údržbou.

Jaká je strategie ukládání (interní, externí, kombinovaná interní a externí podpora)?	

Otázka 4.4 Strategie softwarové podpory (Software Management)

Softwarovou strategií myslíme způsob, jakým repozitář získává a provozuje software nezbytný pro fungování repozitáře. Nejde o rozdíl mezi *„open source“* softwarem a komerčním softwarem, ale spíše o strategii repozitáře v oblastní zajištění provozu softwaru. Nejobvyklejší jsou následující strategie:

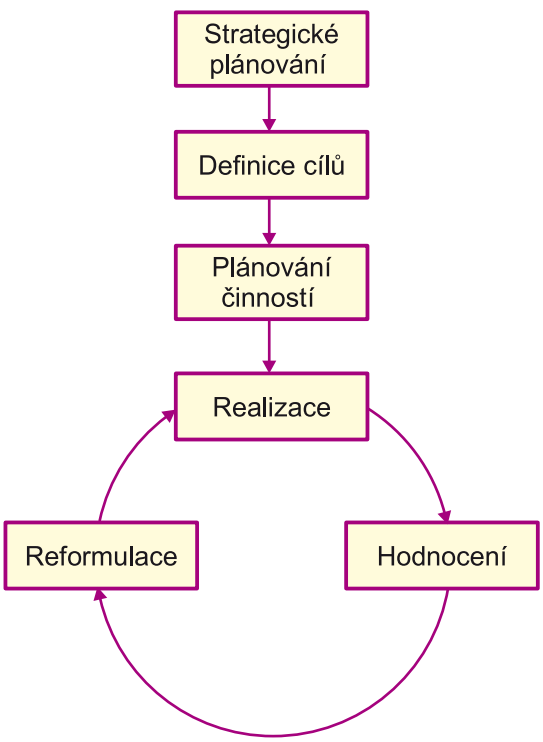
- podpora softwaru dodavatelem
- podpora softwaru někým jiným (třetí stranou)
- vlastní podpora (interní)
- podpora od komunity uživatelů a vývojářů

Jaká je strategie softwarové podpory? (software management)	

4 PLÁNOVACÍ CYKLUS PLATTER

PLATTER plánovací cyklus popisuje formalizovaný soubor kroků, které mají usnadnit definování a specifikaci cílů organizace; implementaci a kritéria k hodnocení toho, nakolik jsou jednotlivé cíle naplňovány.

Je to cyklický proces a v mnoha částech odpovídá analýze rizik DRAMBORA. V následujícím textu poněkud detailněji popíšeme jednotlivé fáze cyklu PLATTER, naznačíme jejich skryté předpoklady a ukážeme, jak spolu jednotlivé fáze vzájemně souvisí.



4.1 Strategické plánování

Strategické plánování je cenným nástrojem k udržení dostatečně široké a na budoucnost orientované organizační perspektivy. Takováto perspektiva je zvláště přínosná v situaci, kdy se jednotlivci obvykle soustředí na bližší a konkrétnější aspekty podnikání. Dobré strategické plánování podporuje nebo ospravedlňuje následná podnikatelská rozhodnutí, je základem pro vytváření detailních plánů, formálním vyjádřením cílů, legitimizací obchodních cílů organizace a velmi napomáhá evaluaci, analytickému modelování a snaze o zlepšování. Strategické plánování souvisí s operačním plánováním, ale liší se od něj. Operační plány se obvykle zabývají mnoha kratšími, explicitnějšími a lépe měřitelnými cíli, kdežto strategické plány jsou širší a obecnější. Operační plánování se odlišuje od strategického managementu a rozhodování, přestože má na ně významný vliv. Ačkoli se strategické plánování zaměřuje na období kolem čtyř let, mělo by podléhat neustálé revizi a mělo by být nejdůležitější činností, která řídí a ovlivňuje vývoj organizace. Obvykle si při strategickém plánování musíme odpovědět alespoň na jednu ze tří základních otázek; odpovědi na ně do značné míry ovlivní další rozhodování:

1. Co děláme?
2. Proč to děláme?
3. Jak v tom můžeme být dobří?

Odpovědi na tyto tři otázky budou zahrnovat mandát repozitáře (nebo nejen mandát, který si stanovuje sama instituce, ale například i legislativní mandát), určí, kdo jsou klíčoví stakeholdeři a jaká mají očekávání, a dále konkretizují podmínky a výkonnost úspěšného repozitáře.

Ke strategickému plánování se obvykle používá postup „*situace > formulace cíle > nalezení cesty k tomuto cíli.*“ Hodnotí se současná situace, následuje formulace cílů a úkolů a nakonec nalezení možných nástrojů k jejich dosažení. Tento postup poskytuje pevnou, ale rozvíjející se základnu pro vývojovou, reprodukční i produkční fázi strategického plánování. Variant na toto základní téma je řada a lze je samozřejmě využít ku prospěchu plánování. Jednou z populárních alternativ je představit si ideální organizační prostředí, porovnat je se současnou provozní realitou a pak komparací rozdílů mezi ideálem a realitou plánovat reorganizaci existujících zdrojů tak, aby se organizace zlepšovala a posouvala se za pomoci vnějších i vnitřních nástrojů k nějakému ideálu.

4.2 Specifikace cílů nebo účelu – operační plánování

Základním vymezením smyslu existence jakéhokoli repozitáře, a vlastně jakékoli formální organizace, je nejobecnější vyjádření základních cílů. Cíle strukturované podle jednotlivých oblastí činnosti jsou implicitně svázány se strategickými plány organizace. Nejužitečnější jsou takové cíle, které odpovídají požadavkům principu „SMART“, tj. jsou konkrétní *(specific)*, měřitelné *(measurable)*, splnitelné *(assigneable, achievable)*, realistické *(realistic)* a mají termín, tj. jsou časově vymezené *(time managed)*. Hodnocení toho, zda bylo či nebylo cílů dosaženo, je možné pouze pro cíle odpovídající principu SMART. Budou-li naše cíle SMART, velmi nám to usnadní plánování a implementaci postupů a metod k jejich dosažení. Strategické plánování by mělo vyjadřovat celkovou filozofii organizace a základní smysl její existence, plánování cílů nebo operační plánování by mělo stanovit praktické postupy, které povedou k dosažení obecně formulovaných cílů.

Cíle by měly brát v úvahu očekávání a požadavky všech hlavních stakeholderů. V případě digitálního repozitáře jsou stakeholdery jejich management, dále ti, kdo repozitář financují, ti, kdo tvoří data, majitelé dat, depozitoři a také uživatelé, kteří mají zájem o obsah uchovávaný v repozitáři. Strategické cíle nebo *„mission statement“* by měly odrážet zájmy všech stakeholderů. V komerčním prostředí budou cíle zahrnovat snahu o růst, zisk, rozvoj infrastruktury a trhu. Ochranné repozitáře mohou být motivovány jinak (růst, profit, vývoj infrastruktury a trhů), ale i jejich motivace by měly být formulovány explicitně. Přesně specifikovat cíle digitální ochrany se může zdát náročné. Je třeba je rozložit na jednotlivé aktivity a role, pak jsou již lépe realizovatelné.

SMART kritéria pro stanovení cílů repozitáře jsou vzájemně kompatibilní a do určité míry logicky svázána. Měřitelné cíle vyžadují konkrétnost a časové termíny. Abychom mohli označit nějaké cíle za realistické či nerealistické, musíme vědět, kdo má odpovědnost za jejich naplnění, a musíme umět přesně kvantitativně vyjádřit jejich splnění. Nestačí ovšem uvažovat o jednotlivých cílech samostatně, je třeba brát v úvahu také vzájemné vztahy mezi nimi. Většina organizací navíc formuluje zároveň velmi různé cíle své činnosti. Často je rozumné jednotlivé cíle hierarchizovat a rozhodnout, které jsou důležitější než jiné. Také můžeme hledat vzájemnou shodu mezi jednotlivými cíli, abychom poznali, zda jsou kompatibilní. Cíle se také mohou lišit pro různé časové periody (krátko-, středně-, dlouhodobé), bezprostřední cíle jsou jiné než operační nebo strategické cíle.

Nejužitečnější je definovat cíle strukturovaně. Jednotlivé cíle spojíme s různými činnostmi organizace, které můžeme rozlišit horizontálně nebo vertikálně. Horizontální přístup strukturuje cíle podle jednotlivých složek organizace, jako jsou zaměstnanci, financování, technologie, právní servis, data, stakeholdeři, plánování a management. Oproti tomu vertikální přístup je zaměřený na procesy nebo funkce. V repozitáři to mohou být řízení, akvizice, archivování, plánování ochrany, ochrana dat, zpřístupňování

dat, což jsou v zásadě jednotlivé procesy popsané v standardním modelu fungování archivu OAIS. V praxi je asi nejvhodnější kombinace obou přístupů, tu například využívá audit rizik DRAMBORA.

Strukturované formulace a rozdělování cílů se musí účastnit řada jednotlivců. Do diskuse by měli přispět zástupci managementu, techničtí specialisté, reprezentanti stakeholderů. K vizualizaci cílů, k odhalení jejich implicitního a měřitelného významu mohou být užitečné manažerské techniky, jako je tvorba myšlenkových map *(mind mapping)*. Pro jednotlivé cíle musí existovat nějaké jednotky, které měří stupeň postupu k cíli. Z pohledu různých repozitářů se tyto jednotky budou velmi lišit v závislosti na kontextech jednotlivých cílů. Cíle spojené s finanční udržitelností mohou používat jako jednotky měnu; cíle spojené se zveřejňováním obsahu repozitáře lze měřit počtem prohlédnutých stránek/objektů na serveru repozitáře; cíle spojené s akvizicí lze měřit počtem digitálních objektů vložených do repozitáře. Složitější může být kvantitativní vyjádření cílů spojených s ochranou dat. Pokud ochranu chápeme jako soubor nějakých důležitých vlastností a jako procedurální a organizační záležitost, lze míru naplnění cílů odhadnout snáze. Příkladem může být migrace obrazového materiálu. Cíle procesu migrace mohou stanovit minimální požadavky na reprodukci barev, zachování rozlišení a technické dostupnosti datových formátů, stanovit časové a jiné limity procesu migrace nebo omezit proces migrace maximálními finančními nebo jinými náklady. Lze využívat jak binární, tak ordinální míry, ovšem použití jemnější stupnice poskytuje podrobnější základ pro hodnocení toho, zda bylo cíle dosaženo. Pokud je používáno více indikátorů, je třeba rozhodnout, které jsou nejdůležitější, abychom usnadnili následné plánování.

Časové parametry by měly být součástí hodnocení postupu u každého cíle a také hodnocení jednotlivců nebo rolí, odpovědných za všechny jednotlivé fáze plánování, implementace a hodnocení.

4.3 Vypracování plánů

Plánování je stavění mostů mezi tím, čeho musí být dosaženo, a tím, co lze reálně dosáhnout. Měli bychom se inspirovat u jiných organizací z podobného prostředí a využít jejich zkušenosti ve svém plánovacím cyklu. Plánování lze také podpořit experimentálním hodnocením, modelováním situací, simulací a tvorbou hypotéz, vždy ovšem s ohledem na strukturu základních strategických cílů. Každé naplánované řešení musí implicitně obsahovat možnost měření výkonu.

Je-li to možné, měli bychom si plánování usnadnit spojováním jednotlivých cílů do skupin. Poměr mezi cíli a plány nemusí být 1:1, ovšem každý cíl musí být nakonec do plánování zahrnut. A samozřejmě bychom neměli plány navrhovat nebo realizovat, pokud nemáme jasno, k jakému organizačnímu cíli směřují.

Tam, kde plánování vede k vzájemnému nesouladu nebo protichůdným krokům, bychom měli upřednostnit důležitější cíle. V některých případech to může vyžadovat dodatečné hierarchické roztřídění cílů a rozhodnutí o tom, které jsou potřebnější.

Plánování procesů *(action planning)* má smysl jen tehdy, chápeme-li ho globálněji, v širší perspektivě organizačních limitů. Mnohé externí vlivy jsou zohledněny už v prvních fázích strategického plánování, především při reflexi současného stavu a analýze možných kontextuálních vlivů. Dosažení cílů plánovaných akcí ovlivňují právní otázky, metodika mateřské organizace, očekávání stakeholderů, dostupnost zdrojů. Je tedy třeba se jimi dostatečně zabývat.

4.4 Realizace plánu, hodnocení/úprava a implementace

V opakujícím se cyklu, který se může rozšířit z fáze plánování a rozvoje repozitáře do fáze plného fungování, jsou tyto vzájemně provázané činnosti základem trvalého zlepšování a rozvoje zralosti repozitáře. Aktivní přístup ke všem třem činnostem prospívá fungování organizace i dosahování jejích cílů. Hodnocení a reformulace plánů by měly následovat nedlouho po implementaci.

Po ukončení úvodní fáze plánování činností by měla co nejdříve následovat první implementace. Jakmile je dosaženo dostatečného stupně rozvoje infrastruktury a jsou-li již použitelné plánované míry výkonnosti, mělo by co nejdříve následovat první hodnocení. Podrobení procesů, procedur a metodik cílenému zkoumání a hodnocení založenému na předem daných kritériích by mělo odhalit nedostatky a možnosti pro zlepšování. Tam, kde jsou důležité časové termíny, by měli hodnotitelé rozhodnout, zda je cílů dosahováno pomocí odhadů a výpočtů.

Reformulace plánů následující po hodnocení by měla být přesnějším a poučenějším opakováním první fáze plánování. Nalezené nedostatky jsou jen dalšími problémy k řešení a plány se mění tak, aby odpovídaly stále lépe poznanému okruhu problémů. Následující fáze by měly přinášet změny stále menšího rozsahu, pokud nejsou změněny původní cíle, čímž je ovšem ukončen jeden plánovací cyklus a zahájen jiný.

5 PLÁNY STRATEGICKÝCH CÍLŮ PLATTERU

PLATTER spojuje soubor plánů strategických cílů s *Deseti základními principy důvěryhodného repozitáře* (viz kapitola 2). Plány strategických cílů se sice úzce váží k *Deseti základním principům*, neplatí však, že jeden plán odpovídá jednomu principu. Repozitáře si samozřejmě mohou vytvářet vlastní plány strategických cílů, ty však logicky budou zahrnovat stejné oblasti jako námi doporučené. Uvědomujeme si, že čtvrtý základní princip týkající se implementace metodiky obsahuje všechny plány strategických cílů, které jako celek reprezentují právě takovouto metodiku.

Plán strategických cílů	Odpovědnost	Odpovídající základní principy
Finanční plán	Zabývá se plánováním, monitorováním a vykazováním financí.	2
Plán řízení lidských zdrojů	Zabývá se získáváním a udržováním souboru dovedností, které jsou relevantní pro správu repozitáře.	2
Datový plán	Specifikuje datové a metadatové objekty, formáty, struktury pro vkládání, uchovávání a zpřístupňování dat a související transformace a mapování dat.	5, 6, 7, 8
Akviziční plán	Zabývá se vztahy s depozitory a dalšími poskytovateli dat.	3, 5
Plán zpřístupňování	Zabývá se vztahy s koncovými uživateli a pravidly pro zpřístupňování.	1, 8
Plán ochrany	Zajišťuje zpřístupňování a použitelnost dokumentů bez ohledu na zastarávání nebo změny technologií.	9
Technický plán	Specifikuje požadavky na hardware, software a síťové systémy.	10
Plán zajištění kontinuity	Zabývá se povinností a zajištěním ochrany dokumentů i po zániku repozitáře.	1
Krizový plán	Reaguje na náhlé změny v prostředí repozitáře.	1, 6

Nyní přejdeme ke konkrétním plánům strategických cílů. *Deset základních principů* převedeme do souboru cílů SMART. Rozbor každého plánu strategických cílů se skládá z obecného úvodu do příslušné oblasti a z popisu cílů daného plánu. Cíle by měly být všeobecné a měly by odpovídat strategické úrovni plánování společné téměř všem repozitářům. Pro každý cíl pak uvedeme několik konkrétních příkladů, tj. realizací všeobecných cílů, jež by mohly vyhovovat nějakému konkrétnímu repozitáři. Příklady je nutno brát jako inspiraci k definování vlastních cílů. Následuje rozbor bodů, na které je třeba se při realizaci všeobecných cílů soustředit.

Při používání PLATTERu bychom neměli opomíjet žádné obecné cíle nebo otázky, které v diskusi vyvstanou. Všechny probírané problémy se samozřejmě netýkají každého repozitáře. Je však důležité, aby si každý repozitář byl všech problematických oblastí vědom. Pokud se repozitář rozhodne, že si v nějaké oblasti žádné cíle nestanoví, měl by takové rozhodnutí explicitně zdůvodnit. Někdy bude zdůvodněním samotný fakt, že daný problém není pro určitý repozitář relevantní. Jiným důvodem může být skutečnost, že by výdaje nebo obtížnost řešení daného problému přesahovaly možný přínos nebo dostupné zdroje. Je třeba si uvědomit, co nás v kontextu našeho repozitáře vede k rozhodnutí nějaké

oblasti se nevěnovat, a je nutné toto rozhodnutí jasně formulovat tak, aby mohlo být zhodnoceno při realizaci plánovacího cyklu PLATTERu.

5.1 Finanční plán

Hlavní hrozbou dlouhodobé ochrany digitálního i jakéhokoliv jiného kulturního dědictví je nedostatek financí. Nejde jen o to, že by repozitář mohl být z důvodu nedostatku peněz zcela uzavřen. Nedostatečné nebo neefektivně využité finanční zdroje mohou způsobit ztrátu použitelnosti dokumentů. Proto musí mít každý digitální repozitář kvalitní projekt podložený finančními zdroji.

Typy repozitářů a způsoby jejich financování se velmi různí, těžko lze mluvit o finančním plánování repozitáře obecně. Důvěryhodnost repozitáře nelze posuzovat na základě schopnosti prokázat garantované zdroje dlouhodobého financování. Málokterý repozitář si totiž může být na mnoho let dopředu jist svými finančními zdroji. Z těchto důvodů se finanční plán zaměřuje na důkladné finanční plánování a monitorování a na vytváření nouzových plánů pro případ finančních problémů. (Tomu, jak se zachovat v případě, pokud by repozitář byl zcela uzavřen kvůli nedostatečnému financování nebo platební neschopnosti, se více věnuje Plán zajištění kontinuity.)

Cíl 1.1 Pravidelně sledovat a revidovat finanční plán

Příklady:

- Účetní uzávěrka a rozpočet se tvoří každý rok k určitému datu. Kontroluje a schvaluje je řídící výbor repozitáře.

Komentář:

Finanční plán je třeba pravidelně aktualizovat. Zdroje financování nejsou obvykle zajištěny na dobu delší než několik let. Proto je třeba monitorovat finanční situaci repozitáře a problémy odhalit dříve, než se rozpočet dostane do schodku.

Cíl 1.2 Udržet financování na takové úrovni, kterou vyžaduje běžný provoz repozitáře

Příklady:

- Udržet financování ve výši 100 % předpokládaného rozpočtu nutného pro běžný chod repozitáře.

Komentář:

Provoz digitálního repozitáře přináší fixní náklady, kterým se nelze vyhnout. Při nedostatečném financování může docházet k porušování závazků repozitáře, a tím k ohrožení jeho důvěryhodnosti. Repozitář tedy musí mít takové příjmy, které zaručí jeho běžný chod. Pokud to není možné, závazky je potřeba pouze přizpůsobit (nikoli zrušit!) tomu, čeho lze dosáhnout s dostupným rozpočtem.

Cíl 1.3 Vytvořit takové nouzové plány pro případ finančních omezení nebo krizí, které dostatečně ochrání důležitá data

Příklady:

- Uzavřít dohodu s jiným repozitářem o tom, že v případě předvídatelných finančních problémů převezme důležitá data.
- Stanovit, jaké služby je třeba udržet i v případě finanční tísně.

Komentář:

Nouzový plán pro případ vážných finančních omezení je jednou z nejcitlivějších oblastí plánování repozitáře. Mnozí takové plány vytvářejí neradi. Bojí se, aby je poskytovatelé financí nebrali jako přiznání, že nynější úroveň financování je zbytečně velkorysá.

Požadavek důvěryhodnosti však znamená, že každý repozitář musí mít plány, jak ochránit nejdůležitější data v případě výrazného úbytku finančních zdrojů. Taková ochrana může nabývat různých podob. Podrobněji o ní pojednáme v Plánu zajištění kontinuity a Krizovém plánu.

Cíl 1.4 Stanovit si a naplňovat takové marketingové plány a plány externí komunikace, které odpovídají potřebám daného repozitáře

Příklady:

- Vytvořit plán externích aktivit ke konkrétnímu datu a revidovat ho každé dva roky.

Komentář:

Komunikace a marketing jsou klíčovými aktivitami repozitáře, zvláště komunikace s depozitory, uživateli, poskytovateli financí a externími spolupracovníky. Repozitář zaměřuje svoje externí aktivity na ty oblasti, které na základě svého současného profilu upřednostňuje. Například repozitáře s jasně vymezenými a dostatečnými zdroji nových dokumentů nebo repozitáře budované pro sklízení veřejně dostupných zdrojů (např. webový archiv) nemusí hledat nové depozitory. Na druhé straně institucionální repozitáře, které počítají s tím, že do nich budou badatelé dobrovolně vkládat své dokumenty, potřebují účinnou komunikační strategii, která vkladatelům vysvětlí, proč je výhodné věnovat čas vkládání dokumentů. Repozitáře by také měly umět využít synergický účinek různých externích aktivit. Například růstem počtu uživatelů nebo významnými mezinárodními kontakty lze argumentovat při hledání dodatečných finančních zdrojů.

5.2 Akviziční plán

Akviziční plán má tři hlavní části: specifikace ukládaných dokumentů, vyjednávání smluv potřebných k jejich získání a vývoj procedur i metod jejich akvizice.

Cíl 2.1 Získat relevantní dokumenty

Příklady:

- Archivovat 90 % národního internetu (např. domena.cz).
- Archivovat 75 % všech článků publikovaných mateřskou institucí.
- Vložit minimálně 10 000 nových obrázků za rok.

Komentář:

Měli bychom určit, kolik dokumentů chce repozitář v daném časovém období získat. Mandát určuje typ repozitáře (archiv, sbírka elektronických publikací, webový archiv atp.), a tím také blíže definuje sbírané dokumenty. Odhad množství/objemu shromažďovaných dokumentů může někdy vyžadovat rozsáhlé analýzy. Například analýzu trhu, která by měla zjistit, jaké dokumenty jsou dostupné; analýzu zájmů jednotlivých stakeholderů, která by měla zjistit přání depozitorů a koncových uživatelů; a analýzu nákladů a výnosů, která by měla odhalit ekonomická omezení. Jindy lze specifikovat shromažďované dokumenty jednoduše až triviálně, například jako všechny dokumenty publikované mateřskou institucí repozitáře. I zde budeme možná potřebovat důkladnou analýzu nebo pilotní projekt k získání realistických čísel.

Cíl 2.2 Vyjednat dohody o uložení

Příklady:

- Vyjednat dohodu o uložení se společností X, která zaručí přístup ke všem svým elektronickým publikacím po dobu nejméně posledních pěti let.
- Vyjednat dohodu o vkládání archivních dokumentů v digitální podobě.

Komentář:

Jaké materiály repozitář od depozitorů získává, by mělo být smluvně upraveno. V některých případech, jako jsou národní knihovny nebo archivy, může být povinnost ukládat dokumenty stanovena zákonem. V jiných případech se dokumenty získávají vyjednáváním s jejich původci nebo dokumenty vytváří sama instituce, která provozuje repozitář, např. digitalizací existující analogové sbírky.

Při specifikaci oblastí, kterých by se měly případné smlouvy o uložení digitálních dokumentů týkat, se můžeme inspirovat manuály Nestor a TRAC:

- množství dokumentů
- způsob dodávání (např. stahování z FTP)
- formáty souborů
- dodávaná metadata
- práva k ochranným opatřením prováděným na dodaných dokumentech (např. migrace, vytváření několikanásobných kopií)
- práva k používání/distribuci dokumentů
- povinnost depozitora upozornit repozitář na veškeré změny týkající se formátů souborů, způsobu dodávání apod.
- případný převod zákonných práv
- odpovídající doba trvání smlouvy

Některé z výše uvedených bodů bude třeba vyjednat a právně dořešit i v případech, kdy jsou dokumenty získávány jako povinné výtisky. Bude například nutné zjistit, zda související zákony umožňují dokumenty pro účely dlouhodobé ochrany kopírovat a/nebo migrovat. Technické podrobnosti týkající se přenosu dokumentů budou obvykle vyžadovat nějakou dohodu. Problematiku zákonných práv musíme mít na zřeteli, i pokud dokumenty generuje sám repozitář. Smí například repozitář digitalizovat svůj knihovní fond z důvodu ochrany? A jak je v tomto případě omezena distribuce digitalizovaných dokumentů?

Problematika akvizice deskriptivních a bibliografických metadat z externích zdrojů je v zásadě totožná s problematikou akvizice primárních dokumentů[14]. Televizní archiv může například získávat deskriptivní a technická metadata o vysílaných pořadech od marketingové agentury specializující se na prodej diváckých statistik a demografických dat potenciálním inzerentům. Repozitář musí vyjednat s poskytovatelem metadat v podstatě stejné body, jako v případě primárních dat.

Cíl 2.3 Získat fyzickou kontrolu nad dokumenty

Příklady:

- Zakoupit a nainstalovat software na sklízení dat.
- Vytvořit pracovní postup pro stahování nových dokumentů z FTP jednou za měsíc.

Komentář:

Tento cíl zohledňuje pouze procesy nezbytné k tomu, aby měl repozitář kontrolu nad aktuálním stavem dokumentů. Zbývající úkoly se týkají vkládání dokumentů do repozitáře a jsou popsány v Datovém plánu.

14 Triviální skutečnost vzhledem k faktu, že „metadata jedné osoby jsou daty jiné osoby“ .

Cíl 2.4 Monitorování akvizice

Příklady:

- Požadované dokumenty musí být možné fyzicky stáhnout z webové stránky společnosti X.
- Minimálně 95 % času televizního vysílání musí být nahráno.

Komentář:

Repozitář musí mít monitorovací systém kontrolující, zda producenti nebo depozitoři požadované dokumenty skutečně poskytují. Tyto systémy mohou být velmi odlišné v závislosti na typu repozitáře. Například institucionální repozitář může od producentů dokumentů vyžadovat porovnání publikací v odborných periodikách a dalších zdrojích se seznamem uložených dokumentů. Internetový archiv může statisticky zhodnotit, kolik procent internetu archivuje. Národní archiv může kontrolovat své depozitory a sledovat, zda plní svou povinnost ukládat povinný archivní výtisk *(legal archive deposit)*.

Cíl 2.5 Zajišťovat aktuálnost smluv o uložení

Příklady:

- Každý rok revidovat smlouvy o uložení.

Komentář:

Smlouvy o uložení musí vždy odpovídat hlavním cílům repozitáře. Repozitář by měl stanovit postupy, jak monitorovat relevanci všech smluv o uložení, a měl by brát v úvahu stejné okruhy problémů, jako při počátečním vyjednávání smluv. Přitom je třeba si uvědomit, že možná bude nutné sjednávat další dohody o uložení s novými poskytovateli.

5.3 Plán řízení lidských zdrojů

Vzhledem k relativně rychlému růstu počtu fungujících repozitářů je obtížné najít kvalifikované zaměstnance. Nejde jen o nedostatek lidí s odpovídající kvalifikací. Systémy repozitářů se velmi rychle vyvíjejí a není jasné, jaké konkrétní zkušenosti a kvalifikace jsou pro práci v repozitáři nezbytné nebo relevantní. Dosud neexistují povolání specializovaná na práci s repozitáři. Proto se zatím zaměstnanci hledají v jiných profesních skupinách – mezi archiváři, knihovníky, IT specialisty, administrátory atd. Tato poměrně nová oblast pracovních příležitostí přináší další problém: neexistuje dosud jasný model kariérního postupu, což snižuje šance udržet zkušené zaměstnance. Repozitáře se společně snaží institucionalizovat zaměstnání v repozitáři a vytvářejí národně a mezinárodně uznávané akreditace a kariérní řády. To je jistě důležité. Jelikož však dosud žádné všeobecně uznávané standardy neexistují, musí jednotlivé repozitáře převzít odpovědnost za definice pracovní náplně jednotlivých zaměstnaneckých pozic a specifikaci kariérního postupu. Jen tak budou moci zaměstnat a udržet si talentované a kompetentní zaměstnance.

Cíl 3.1 Definovat zaměstnanecké pozice, odpovědnosti a pravomoci pracovníků repozitáře

Příklady:

- Podrobně specifikovat zaměstnanecké pozice.
- Jednou za dva roky zaměstnanecké pozice revidovat.

Komentář:

Provoz repozitáře vyžaduje znalosti z mnoha různých oborů. Přesné vymezení odpovědnosti umožní zaměstnancům různého vzdělání a zkušeností plně se soustředit na svou odbornou oblast, a zvyšovat tak kvalitu své práce.

Jasné přidělení odpovědnosti také zjednodušuje organizační strukturu. Pokud například máme pouze jednu pozici pro správu serverů, každý bude okamžitě vědět, na koho se má v případě potřeby obrátit.

Je-li na jasně stanovené odpovědnosti vázán také rozpočet, zaměstnanci jsou lépe motivováni a pomáhá to vybudovat kariérní řád. Vedoucí repozitáře se pak mohou více soustředit na celkovou strategii organizace a nemusí se zdržovat schvalováním výdajů na menší položky.

Cíl 3.2 Získat a udržet zaměstnance pro práci na specifických pozicích

Příklady:

- V rámci repozitáře by měla být ke konkrétnímu datu vytvořena a obsazena pozice manažera a administrátora.

Komentář:

Má-li repozitář stanovené jednotlivé pozice a odpovědnosti, musí také zajistit, aby na ně získal kvalifikované zaměstnance. Nedostatek lidských zdrojů může časem vést k porušování závazků repozitáře, což by se nepříznivě projevilo na jeho důvěryhodnosti.

V mnoha případech bude se získáváním zaměstnanců repozitáře také souviset jednání o využití těch, kteří v současnosti již pracují v mateřské organizaci.

Cíl 3.3 Rozvíjet kvalifikaci zaměstnanců

Příklady:

- Všichni zaměstnanci se jednou ročně zúčastní hodnocení osobního rozvoje. Zároveň si stanoví a zhodnotí cíle svého individuálního růstu.
- Alespoň jednou ročně je nutné financovat účast veškerých zúčastněných zaměstnanců na relevantním mezinárodním semináři nebo konferenci.

Komentář:

Nejvyšší prioritou repozitáře je neustálé zvyšování kvalifikace zaměstnanců. Překotný růst v oblasti repozitářů vede k rychlému zastarávání jejich dovedností. Repozitář by se měl starat o další vzdělávání svých pracovníků.

Standardy pro činnosti repozitáře, pro nakládání s riziky a pro způsoby zprostředkování dat uživatelům se dosud vyvíjejí. Pracovníci starající se o repozitář by se měli účastnit seminářů, uživatelských skupin a konferencí, aby věděli o aktuálním vývoji v oblasti provozování repozitářů.

Repozitáře by se také měly vyvarovat vytváření příliš specializovaných pozic. Když zaměstnávají nenahraditelné odborníky, vystavují se riziku, pokud tito specialisté odejdou. Repozitář by měl proto podporovat sdílení znalostí mezi zaměstnanci, a tím snižovat důsledky takovýchto situací.

5.4 Plán zpřístupňování

Distribuce obsahu je jedním z nejdůležitějších a nejviditelnějších výstupů téměř všech repozitářů. Vedle archivní funkce je jedním z hlavních důvodů zakládání repozitářů (univerzitních repozitářů, archivů elektronických publikací atp.) právě distribuce a zpřístupňování dokumentů.

Existují nicméně repozitáře, které nezpřístupňují žádné dokumenty, jejich hlavní funkcí je pouze bezpečné uchovávání dat (tzv. *„dark archives“*). Takové repozitáře se samozřejmě nemusí příliš zabývat problémy souvisejícími s distribucí dokumentů. Ovšem i tyto repozitáře musí počítat s nějakými uživateli, třeba s budoucími generacemi historiků. Při plánování repozitáře je třeba zohlednit alespoň minimálně potřeby budoucích uživatelů. Přinejmenším popisná metadata pomohou uživatelům orientovat se v archivovaných dokumentech a rozumět jejich obsahu.

Z hlediska zpřístupňování obsahu repozitáře je třeba věnovat pozornost především následujícím otázkám:

1. Komu chceme nebo potřebujeme poskytovat data z našeho repozitáře?
2. Jaké jsou podmínky a omezení, v rámci kterých bude umožněn přístup?
3. Jaké technické prostředky, opatření a zařízení potřebujeme mít?

Cíl 4.1 Formulovat, udržovat a aktualizovat programové prohlášení, které bude odpovídat mandátu repozitáře

Základní funkcí každého repozitáře musí být ochrana dokumentů pro budoucí generace, proto se domníváme, že diskuse o programovém prohlášení repozitáře patří právě do rozpravy o plánování distribuce.

Příklady:

- Řídící výbor repozitáře schválí programové prohlášení ke konkrétnímu datu.
- Každé dva roky zkontrolovat, zda repozitář plní své programové prohlášení.
- Každých pět let revidovat relevanci programového prohlášení.

Komentář:

Programové prohlášení obvykle vychází z mandátu repozitáře. Rozdíl mezi programovým prohlášením a mandátem spočívá v tom, že mandát je většinou přidělen vnějším subjektem odpovědným za založení repozitáře, například ministerstvem nebo mateřskou organizací. Programové prohlášení formuluje sám repozitář (instituce), mělo by obsahovat závazek repozitáře chránit soubor dat, dokumentů nebo znalostí v zájmu konkrétní skupiny koncových uživatelů.

Programové prohlášení je velmi důležité, protože pomáhá naplňovat mandát repozitáře a vymezuje vztahy s externími subjekty. Programové prohlášení by mělo být živým dokumentem. Repozitář by jej měl pravidelně revidovat, hodnotit a zajistit tak nepřetržitý vývoj.

Programové prohlášení by například mělo:

- definovat cíle repozitáře,
- obsahovat závazek dlouhodobého uchovávání, správy a zpřístupňování digitálních informací definované komunitě uživatelů,
- obsahovat závazek udržovaní vztahů vzájemné důvěry mezi repozitářem a jeho stakeholdery, např. prostřednictvím auditu a certifikace.

Cíl 4.2 Definovat komunitu/y uživatelů repozitáře, rozumět jejich potřebám a dovednostem

Příklady:

- Definovat cílovou skupinu konečných uživatelů k určitému dni.
- Revidovat potřeby uživatelů jednou za dva roky.
- Vytvořit kontaktní skupinu, která se bude s reprezentanty uživatelů každého půl roku setkávat.

Komentář:

Na webové stránce repozitáře by měla být cílová skupina uživatelů repozitáře popsána. Uživatelé by měli mít jasný přehled o možnostech a limitech získávání dokumentů. Repozitář by nikdy neměl ztratit přehled o potřebách uživatelů, měl by tyto potřeby průběžně monitorovat a uživatelům se přizpůsobovat.

Při definování uživatelské komunity je třeba si ujasnit následující:

- Kdo je cílovou skupinou?
- Jak velká je tato skupina a jak bude růst?
- Jak různorodá je tato skupina (z hlediska věku, profese, prostředí atd.)?
- Jaká je znalostní základna této skupiny?
- Jakou úroveň služeb očekává?

Repozitář by měl mít nějaké nástroje pro sledování změn potřeb uživatelů. Plán zpřístupňování by měl stanovit, jak bude repozitář zjišťovat změny v očekávání a potřebách uživatelů, zda například použije nějaké dotazníkové šetření, formální hodnocení, workshop nebo individuální konzultace apod.

Cíl 4.3 Formulovat a implementovat politiku zpřístupňování obsahu repozitáře

Příklady:

- Všechny dokumenty jsou volně dostupné *(Open Access)*.
- Všechny dokumenty jsou dostupné registrovaným uživatelům.
- Dokumenty jsou prezenčně dostupné všem uživatelům, absenčně pouze vybraným badatelům.
- Revidovat jednou ročně, zda repozitář funguje v souladu s politikou zpřístupňování obsahu.
- Politiku zpřístupňování obsahu revidovat jednou za dva roky.
- Vytvořit k určitému dni metodiku pro zacházení s návrhy nebo stížnostmi, které se týkají politiky zpřístupňování obsahu.

Komentář:

Přístup repozitářů k politice zpřístupňování obsahu může být velmi různorodý. Většina repozitářů se snaží distribuovat své dokumenty co možná nejširšímu okruhu uživatelů. Problémem je zjistit, co je možné a co není. Omezení mohou pocházet z těchto zdrojů:

- Autorské právo
- Zákon o šíření soukromých nebo obchodně citlivých dat
- Národní bezpečnost
- Zákony týkající se pomluv, obscénností, projevů nenávisti, nactiutrhání apod.
- Specifická smluvní omezení, která stanovují dohody o uložení

Omezení mohou jednotlivé skupiny uživatelů různě ovlivňovat. Některé dokumenty mohou být například dostupné pouze badatelům, ale veřejnosti nikoli. Politika zpřístupňování obsahu by také měla definovat práva zaměstnanců repozitáře nebo technickou podporu. Metadata by měla zohledňovat omezení politiky zpřístupňování obsahu např. tak, aby některé dokumenty byly v souladu se zákonem dostupné pouze dospělým.

Politika zpřístupňování obsahu by se také měla zabývat autorizací, ověřováním (autentikací) a způsobem přihlašování uživatelů. Zavádění systému kontroly přístupu, který implementuje politiku zpřístupňování obsahu, je předmětem Technického plánu. Součástí Plánu zpřístupňování by však také měla být kontrola toho, zda je tato implementace v souladu s cíli zpřístupňování.

Cíl 4.4 Specifikovat a realizovat technologické požadavky distribuce a zpřístupňování

Příklady:

- Stanovit a implementovat minimální požadavky na metadata, která pomohou cílové skupině vyhledat požadované dokumenty (připravit k určitému dni).
- Implementovat takové vyhledávací nástroje, které dokáží vyhledat dokumenty podle názvu nebo autora (připravit k určitému dni).
- Všechny dokumenty zpřístupnit nejpozději do tří dnů po akvizici.

Komentář:

Zatímco se cíl 4.3 zabýval omezením přístupu k dokumentům, cíl 4.4 se zabývá vytyčením cílů pro podporu zpřístupňování. Zásadní význam zde mají deskriptivní a bibliografická metadata, jelikož uživatelům umožňují najít a získat požadovaný dokument. Repozitář by si měl vytyčit cíle pro vyhledávání a elektronickou distribuci konkrétních digitálních objektů a přitom brát v potaz předpokládané sys-

témové požadavky a náklady. Technická podoba distribuovaných dokumentů neboli distribuovaného informačního balíčku (DIP) spadá pod odpovědnost Datového plánu. Dalšími uvažovanými oblastmi jsou přihlašování a používání DRM (*Data Right Management* – správa práv k datům) pro distribuované dokumenty.

5.5 Technický plán

Pro ochranu digitálních dat je samozřejmě rozhodující IT infrastruktura. Dokumenty jsou ohroženy rozmanitými problémy. Posláním Technického plánu je řešení rizik, která hrozí datům a systému, který data uchovává. A také rizik, která hrozí systémům, jež poskytují služby koncovým uživatelům. Ačkoli je nemožné ochránit všechny složky systému správy a zpřístupňování digitálních dat dokonale, technický plán by měl vyjmenovat co možná nejvíce možných rizik a neměl by ignorovat ta, pro která nemá momentálně řešení. Technické požadavky se běžně dělí na softwarové, hardwarové a síťové. Pro potřeby managementu cílů je ale lepší uvažovat o všech rizicích obecně bez ohledu na tyto kategorie. V PLATTERu si vytyčujeme cíle ve třech hlavních oblastech technické infrastruktury – rozsah, bezpečnost a služby.

Cíl 5.1 IT infrastruktura si musí umět poradit s takovým rozsahem ukládání, zpracování a přenosu dat, který odpovídá potřebám daného repozitáře

Příklady:

- Repozitář musí mít externí síťové připojení s minimální rychlostí x Mb/s.
- Rychlost vnitřní sítě se musí pohybovat v gigabitech.
- Repozitář musí mít takovou výpočetní kapacitu, že dokáže denně převést x GB videa do formátu MPEG.
- Repozitář musí mít pro externí služby škálovatelné serverové řešení.
- Každý rok musí být provedena kontrola, zda mají systémy pro chod repozitáře stále dostatečnou kapacitu.

Komentář:

Obecně vzato musí IT infrastruktura udržovat služby na takové úrovni, kterou vyžadují závazky repozitáře. Přitom je nutno brát v potaz jak změny v samotném repozitáři, tak ve vnějším prostředí, např. zvyšující se provoz webu nebo nové bezpečnostní hrozby. Přesné schéma IT infrastruktury repozitáře se do značné míry bude odvíjet od jeho závazků, nicméně všechny repozitáře se musí zabývat následujícími problémy:

- Jsou hardwarové, softwarové a síťové systémy dostatečné pro množství dat, které přijímají?

Repozitář si musí vytvořit takovou IT infrastrukturu, která bude vyhovovat požadavkům a zátěži, které bude přijímat. K prodloužení doby provozuschopnosti by měly být využity záložní systémy, které sdílejí zátěž.

- Provoz na dostatečně podporovaných operačních systémech a dalším softwaru základní infrastruktury.

Problém není v tom, zda software má, nebo nemá otevřený kód. Mnohem důležitější je jeho náležitá podpora. Ta může být zprostředkována formální dohodou o poskytování podpory, kontaktem se skupinami uživatelů vývojářů nebo vývojem a správou v rámci samotného repozitáře. Důležité je sledovat, zda je úroveň podpory dostatečná.

Usnadnit podporu a snížit její cenu lze používáním systémů jiných repozitářů, účastí v uživatelských skupinách a zajištěním vysoké úrovně školení zaměstnanců.

- Analyzovat softwarové nebo hardwarové systémy, kterým se již nedostává dostatečné podpory, a proto musí být nahrazeny.

Repozitář musí stanovit metody, jak sledovat stáří hardwaru a softwaru a jak zjistit, zdali tento má, nebo již nemá odpovídající podporu.

- O všech změnách v IT infrastruktuře si vést záznamy.

Důležitým krokem na dlouhé cestě k získání důvěryhodnosti je uchovávání záznamů o všech změnách a procesech v IT infrastruktuře. Nedostatky v dokumentaci změn a procesů spolu s vysokou fluktuací zaměstnanců představují vážné hrozby pro bezpečnost a integritu dat. Pokud máme všechny změny v systému zdokumentované, můžeme také rychleji měnit lidské zdroje, neboť noví zaměstnanci se nemusejí opírat o paměť jiných zaměstnanců, ale mohou se podívat do záznamů.

Cíl 5.2 Infrastruktura IT musí garantovat integritu a bezpečnost uložených dat

Příklady:

- Pro snadnější obnovu dat uchovávat tyto v systému RAID.
- Páskové zálohy všech dat ukládat jednou týdně (denně) mimo místo provozu.
- Každé tři měsíce generovat kontrolní součty a porovnávat je s předchozími hodnotami, aby se zjistily změny v datech.
- Komunikaci mezi servery filtrovat firewallem a fyzický přístup k serverům povolit jen vybraným zaměstnancům.
- Každých X let podrobit repozitář bezpečnostnímu auditu.

Komentář:

Bezpečnost a integrita ve všech svých aspektech jsou základními podmínkami správného fungování digitálního repozitáře. Data nesmí být měněna bez autorizace a musí být prokazatelné, že jsou stále stejná. Repozitář musí zdokumentovat hrozby, proti kterým se chrání, a opatření, která k tomu používá.

Je třeba vzít v potaz následující čtyři aspekty:

- Ochránit digitální data před neautorizovaným užitím (prohlížením apod.).

Mnoho repozitářů uchovává data, která jsou obchodně, politicky nebo jinak citlivá. Pokud se k takovým datům dostanou neautorizovaní uživatelé, následky mohou být někdy značné. Sem patří případy, kdy si zaměstnanci berou neoprávněně data domů, kde jsou mimo ochranu repozitáře a mohou být zpřístupněna nebo odcizena. Otázky politiky zpřístupňování obsahu jsou probrány v Plánu zpřístupňování. Technologické implementace musí podporovat politiku zpřístupňování dat daného repozitáře.

- Ochránit digitální data před změnou neautorizovanými uživateli.

Data jsou důvěryhodná, pokud je každá jejich změna zdokumentována. Mohou-li uživatelé data změnit, aniž by tato změna byla zaznamenána, nelze již data považovat za spolehlivá. Důležitým krokem k získání důvěryhodnosti repozitáře je důsledné zaznamenávání změn v datech, a to i v případě, že změny není možné vrátit.

- Zabránit zničení digitálních dat událostmi spadajícími mimo působnost repozitáře.

Přírodní katastrofy mohou IT systémy snadno zničit, pokud nejsou vybudovány tak, aby jim odolaly. Stejně tak může repozitář přijít o své systémy i zaměstnance v důsledku politické nestability nebo války a následného rabování. Dokonce i tak prostá závada jako výpadek proudu může poškodit citlivé systémy. Tato rizika jsou probrána v Krizovém plánu.

- Zabránit zničení digitálních dat událostmi spadajícími do působnosti repozitáře.

Viry a trojské koně mohou velmi snadno ohrozit a poškodit nezabezpečené IT systémy. Velmi nebezpečný může být tzv. *ransom-ware*, software, který data zašifruje a pak za výkupné nabídne jejich zpětné rozšifrování. Takováto rizika a záškodnické aktivity mohou způsobovat i sami zaměstnanci. S tím by měla ochranná opatření počítat a zajistit např. to, aby žádný zaměstnanec neměl přístup ke všem on-line datům i záložním kopiím současně.

Cíl 5.3 Infrastruktura IT musí garantovat dostupnost daných služeb pro uživatele

Příklady:

- Jednou za tři měsíce provést kontrolu statistik používání všech služeb.
- Zajistit provozuschopnost vyhledávacích a zpřístupňovacích služeb ve výši 99 % (měřeno čtvrtletním průměrem).
- Informace o uživatelích a jejich přihlášení uchovávat v bezpečném systému, který je oddělen od systému poskytujícího služby.
- Rozdělit výkon služeb do více virtuálních strojů, čímž se bude dynamicky zvyšovat nebo snižovat dostupnost zdrojů a usnadní se obnova nefunkční služby.
- Vlastnit další zařízení se záložními kopiemi, které může v případě lokální katastrofy pokračovat v provozu služeb.
- Uzavřít dohodu s poskytovatelem internetových služeb o opatřeních pro případy ztráty konektivity nebo DDoS útoků (*Distribuované odmítnutí služby* – přehlcení služby požadavky).

Komentář:

Poruchy v oblasti služeb se nepříznivě odráží na důvěryhodnosti repozitáře, protože služby reprezentují závazky repozitáře.

Využívání služeb repozitáře je pro mnoho uživatelů jedinou interakcí s repozitářem. Když tyto služby nebudou k dispozici nebo nebudou dobře fungovat, nepříznivě to ovlivní pohled uživatelů na celý repozitář.

K dosažení tohoto cíle je třeba zvážit následující okruhy:

- Zjistit, o jaké služby již není zájem a lze je zrušit a po jakých službách je naopak velká poptávka.

S tím, jak se repozitář rozvíjí a mění se uživatelská komunita, nemusí již být o některé služby zájem, tudíž by měly být zrušeny. Sem patří staré vyhledávací systémy, zastaralé formáty zobrazení a poskytování dat na starých formátech médií.

Uživatelé by také mohli požadovat nové způsoby prezentace dat.

- Zabránit zpřístupnění informací o uživatelích bez řádné autorizace.

Nejcitlivější bývají informace o tom, k jakým datům dotyčný uživatel přistupoval nebo jaké osobní informace vyplňoval při registraci. I když takovéto informace může repozitář uchovávat, musí současně poskytnout záruky, že nejsou přístupné jiným uživatelům z důvodu špatně zabezpečených služeb.

- Zabránit přerušení služeb událostmi spadajícími do působnosti repozitáře.

Z hlediska uživatele by neměla mít porucha jednoho systému repozitáře vliv na jeho služby. V případě takové poruchy by se mělo využít základních záložních služeb typu „*hot-swappable systems*“ (systémy pro náhradu hardwaru za chodu).

- Zabránit přerušení služeb náhodnými událostmi spadajícími mimo působnost repozitáře.

Z pohledu uživatele by měly služby fungovat i v případě negativních vlivů okolního prostředí. Mezi takové vlivy patří výpadek proudu nebo internetového připojení nebo také závažnější záležitosti jako záplavy či požár. Mezi možnosti, jak se takovým vlivům bránit, patří pořízení si dalšího zařízení, které bude umístěno mimo místo provozu primárního systému, a v případě havárie tak zajistí pokračování provozu služeb. Vhodné je také opatřit si několik internetových připojení najednou a zajistit si nouzové napájení. Jednotlivé repozitáře budou muset provést rozbor nákladů a přínosu takovýchto záložních služeb v porovnání s pravděpodobností jejich výskytu.

Negativní vliv na provozuschopnost nebo důvěryhodnost služby mohou mít útoky DDoS, vedené hackery s cílem zničit server. Před takovými útoky musí repozitář své systémy chránit.

5.6 Datový plán

Datový plán popisuje jednak formáty dat a metadat, které užívá repozitář, jednak konverze užívané během vkládání a zpřístupňování. Plán dále popisuje strategie, jak sledovat vhodnost vybraných formátů. V této části práce budeme používat obecnou terminologii, zavedenou v referenčním rámci OAIS:

SIP: Vstupní informační balíček
AIP: Archivní informační balíček
DIP: Distribuovaný informační balíček

Tyto termíny popisují stadia transformací, kterými procházejí data a metadata v repozitáři. Předpokládanou cestu dat repozitářem lze popsat takto:

1. Poskytovatel zakóduje svá data do takového formátu vstupního balíčku (SIP), který repozitář akceptuje.
2. Repozitář přijme balíčky SIP a transformuje je pro archivaci (balíčky AIP).
3. Uživatel repozitáře požaduje určitá data a repozitář mu je dodá transformovaná do patřičného formátu pro distribuci (DIP).

Pro každou fázi musí být definovány formáty dat a metadat a také způsoby převodů mezi nimi.

Cíl 6.1 Určit, jaké formáty digitálních objektů bude repozitář akceptovat (SIP)

Příklady:

- Repozitář internetového archivu bude stahovat a uchovávat stránky v původní podobě s hlavičkami http, které poslouží jako dodatečný zdroj metadat.
- Archiv elektronických publikací bude akceptovat soubory ve formátu PDF vložené prostřednictvím webového formuláře s doprovodnými metadaty.

Komentář:

Pro ochranu digitálních dokumentů je klíčové specifikovat formáty, v kterých bude repozitář přijímat digitální objekty. Je snadné vybudovat repozitář, který uchovává pouze digitální bity, bity samy o sobě ovšem mohou být velmi brzy k ničemu. Proto musí repozitář specifikovat, které formáty datových souborů akceptuje a jaká metadata (a jejich formát) by měla doprovázet každý digitální objekt.

Problémy s formáty souborů někdy působí naprosté detaily. Některé formáty konvertovat nelze: formáty (potenciálně) obsahující šifrování nebo vložené objekty a obrázky či komerčně chráněné formáty. Tyto soubory lze uchovávat, ale jakmile vlivem technologického vývoje zastarají, bude jejich obsah ztracen. V takovém případě nelze tedy garantovat dlouhodobou ochranu.

Repozitář by měl mít návod pro poskytovatele dat, popisující jaká data, metadata a dodatečné informace o formě reprezentace *(representation information)* mají obsahovat balíčky SIP a jak mají být zabaleny. Repozitář by také měl kontrolovat kompletnost a bezchybnost vkládaných balíčků SIP a měl by mít zavedené nějaké procedury pro nakládání s neplatnými nebo nekompletními balíčky SIP. V některých případech (např. webový archiv) stačí aplikovat pravidlo „stahovat vše“, v jiných případech bude třeba mnohem přísnějších kritérií.

Repozitář bude pravděpodobně pracovat s několika druhy digitálních objektů. Pro každý takový druh by měly být definovány požadavky na balíčky SIP a ověřovací metody.

Cíl 6.1.1 Specifikovat zdroje a formáty bibliografických a popisných metadat pro SIP

Příklady:

- Ručně opsat metadata (např. autor, název, klíčové slovo apod.) z elektronických publikací.
- Archiv televizního vysílání získá metadata z programu, který dodá tisková kancelář.
- Deskriptivní metadata pro lékařské snímky zapíše lékař během vkládání snímků.

Komentář:

Bibliografická a deskriptivní metadata specifikují, čím objekt je, co obsahuje a v jakém kontextu byl vytvořen. Tyto základní informace jsou potřebné k tomu, aby byl objekt nalezen (např. vyhledávačem) a interpretován. Metadata mohou být v závislosti na typu repozitáře získávána z různých zdrojů. Výsledná metadata budou v případě většiny repozitářů kompromisem mezi potřebami a přáními uživatelů a náklady spojenými se získáním a zajištěním kvality metadat.

Cíl 6.1.2 Specifikovat technická metadata pro SIP

Příklady:

- Ukládá se pouze velikost souboru a kontrolní součet *(checksum).*
- Informace o formátu získané automaticky standardními nástroji.

Komentář:

Technická metadata popisují vztah mezi digitálním objektem a zdrojem jeho obsahu a formu digitálního objektu samotného (např. formáty dat a provedené konverze). Obecně platí, že minimálním požadavkem dlouhodobé ochrany bitů je ukládání kontrolních součtů pro všechny objekty. Technická metadata mohou být však mnohem bohatší. Mohou například obsahovat podrobné informace o formátech, o softwaru, na kterém byla digitální data vytvořena, atd.

Určité nástroje (např. *JHOVE* nebo *New Zealand Metadata Extractor*) dokáží automaticky extrahovat technická metadata i informace o formátech během procesu vkládání[15].

Cíl 6.2 Specifikovat formát dat a obsah metadat pro archivaci digitálních objektů (AIP)

Příklady:

- Video soubory uložit ve formátu MPEG kompatibilním s DVD.
- Textové soubory uložit ve formátu PDF/A.

15 Více nástrojů je k dispozici na adrese Kongresové knihovny: http://www.loc.gov/standards/premis/tools.html

Komentář:

Pro archivované digitální objekty musí repozitář stanovit formáty souborů a potřebná metadata. V případě balíčků SIP je potřeba hledat rovnováhu mezi běžně užívanými formáty a formáty vhodnými pro archivaci, v případě archivních balíčků AIP to není třeba. Stačí specifikovat, jaké typy dat a metadat je repozitář schopen dlouhodobě ochraňovat. Takové rozhodnutí bude asi také výsledkem rozboru nákladů a přínosů, možná volbou mezi „nekomprimovaným" *(raw)* formátem, bezztrátovou a ztrátovou kompresí. Dalšími zvažovanými faktory budou například dodatečná rizika, která se váží k patentovaným formátům.

Repozitáře by se měly smířit s tím, že se v některých případech není možné vyhnout konverzím. A to i přesto, že počítačová věda nemůže poskytnout matematický důkaz o kvalitě konverze. Repozitář by proto měl disponovat metodou, jak ověřit kompletnost a bezchybnost nově generovaných balíčků AIP i SIP.

Cíl 6.2.1 Specifikovat metadata pro AIP

Příklady:

- Dublin Core XML uložit spolu se zdrojovými daty.

Komentář:

Metadata z balíčků SIP by se měla uchovávat a upřesňovat, nikoliv měnit. Metadatový soubor by měl být čitelný jak lidmi, tak stroji, např. XML.

Kromě metadat z balíčků SIP by balíčky AIP měly obsahovat alespoň následující metadata:

- Záznamy o opatřeních provedených v zájmu dlouhodobé ochrany na digitálních objektech a dalších zásazích do digitálních objektů.
- Záznamy o zákonných a administrativních právech vztahujících se k digitálnímu objektu.
- Jedinečný identifikátor.

Do balíčku AIP je třeba zakódovat, jaké organizační útvary zodpovídají za digitální objekt a jaká zákonná práva se k objektu vztahují. Tyto informace by měla obsahovat smlouva o uložení, v rámci které byl digitální objekt získán (viz Akviziční plán).

Pro určitý typ repozitářů budou relevantní ještě další metadata, která by se měla také zaznamenat, např. strukturální vztahy mezi balíčky AIP.

Jedinečný identifikátor musí být jednoznačný alespoň v rámci daného repozitáře. Repozitáře také mohou využít služeb globálně jedinečných identifikátorů.

Cíl 6.3 Specifikovat formáty dat pro digitální objekty distribuované uživatelům (DIP)

Příklady:

- Na požádání převádět audio soubory z formátu WAV do formátu MP3.

Komentář:

Repozitář musí ve spolupráci s uživateli stanovit vhodné formáty pro distribuci digitálních objektů. Balíčky DIP se mohou v závislosti na typu uživatelů velmi lišit. Od jednoduchých – například metadata kódovaná podle normy OAI nebo Dublin Core obsahující odkazy na nekomprimované datové soubory v balíčcích AIP, až po webovou stránku s podrobnými metadaty a soubory dat konvertovanými do moderních (ztrátových) formátů nebo i extrémnější příklady.

Pro každé schéma balíčku AIP musí existovat několik schémat balíčků DIP. V závislosti na kontextu, v kterém je distribuován, může mít tentýž balíček AIP k sobě připojených několik schémat balíčku DIP. A některé balíčky AIP nemusí mít žádná schémata DIP, což de facto znamená, že tyto digitální objekty uživatelům distribuovány nejsou.

Schémata balíčku DIP se časem mohou, a pravděpodobně budou, měnit s tím, jak se mění schopnosti a nároky uživatelů. Změna schématu balíčku DIP by mohla vést ke změnám AIP, ovšem neměla by je vyžadovat.

Kontrola kompletnosti a bezchybnosti balíčků DIP před tím, než je přijmou uživatelé, může být užitečná, není však nezbytně nutná, jako je tomu v případě balíčků AIP a SIP.

Cíl 6.3.1 Specifikovat metadata pro DIP

Příklady:

- Metadata pro soubory MP3 (viz například Cíl 6.3) zakódovat do ID3 tagů.

Komentář:

Prezentaci metadat je třeba věnovat zvláštní pozornost. Pro uživatele jsou někdy zajímavější než samotná data. Dostupná by měla být veškerá metadata z balíčku AIP včetně metadat z balíčku SIP. Je třeba rozhodnout, jaká metadata jsou relevantní a jaká nadbytečná. Mnoho vyhledávačů nedokáže přečíst samotné datové soubory a hledají balíčky DIP pouze pomocí jejich metadat. Proto mají metadata pro uživatele obrovský význam.

Cíl 6.4 Specifikovat transformaci SIP do AIP

Příklady:

- Balíčkem SIP je zvukové CD. Je potřeba konvertovat stopy z CD do souborů formátu WAV, vyhledat informace o CD v databázi CDDB a uložit je jako dokument XML. Zároveň obal CD naskenovat a uložit ve formátu TIFF.

Komentář:

Repozitář musí jasně definovat postupy vytváření balíčků AIP z balíčků SIP. To znamená popsat, jak se z balíčku SIP získávají metadata, jak se k nim přidávají informace vyplývající ze smlouvy o uložení a další relevantní specifikace, jak jsou pak metadata kódována do formátu pro ukládání metadat balíčku AIP. Někdy bude potřeba digitální data z balíčku SIP přetransformovat do jiného formátu, který bude pro archivaci vhodnější. Metadata o takovýchto změnách by pak měla být rovněž uložena do balíčku AIP.

Dále je nutné vygenerovat jednoznačný identifikátor balíčku AIP a zakódovat všechna další zde neuvedená metadata.

Pak je nutné ověřit správnost a úplnost balíčku AIP a také, zda je každý, ať už používaný, nebo vyřazený balíček náležitě zdokumentován.

Cíl 6.5 Specifikovat transformaci AIP do DIP

Příklady:

- Soubory WAV získané z CD se převedou do formátu MP3. Metadata jsou zakódována jako ID3 tagy do každého souboru. Soubor s naskenovaným obalem se převede do JPEG a zabalí se spolu se soubory MP3 do archivu ve formátu ZIP a takto je nabízen uživateli ke stažení.

Komentář:

Repozitář by si měl stanovit metody transformací, které bude používat pro konverzi balíčku AIP do balíčku DIP.

Jedním ze základních požadavků těchto konverzí je, aby balíčky DIP byly autentickými kopiemi obsahů původních balíčků SIP, nebo aby to byly objekty, z kterých je možné odvodit originály bez ohledu na kódovací formáty.

Tento postup lze snadno vysvětlit, ale v praxi může být technicky složitý. Může sem patřit konverze formátu souboru, opětovné zakódování metadat, které mohou být požadovány v reálném čase během žádosti o dodání.

V ideálním případě by měl repozitář umět při nástupu nových technologií změnit tento postup, aniž by přitom měnil formát balíčku AIP nebo DIP.

5.7 Plán zajištění kontinuity

Repozitáře nejsou jen hardwarem a softwarem, jsou organizacemi, a jako každá organizace jednou, možná relativně zakrátko, přestanou existovat. Citace z knihy Gregory Benforda „V hloubi času" hezky ilustruje, co repozitáře čeká:

„Budoucnost předvídáme na základě analýzy minulosti, hledáním dlouhodobých trendů. Stěží se tak něco dozvíme o vzdálené budoucnosti přesahující tisíc let.

Před zhruba dvěma stoletími bylo to, čemu dnes říkáme Spojené státy, součástí britského koloniálního impéria. V Evropě té doby možná najdeme něco, co se podobalo dnešnímu světu, některé státy přežily takhle dlouho. Pro takovéto časové období je extrapolace vhodná alespoň k předvídání rozsahu a směru toho, co se může stát.

Vrátíme-li se o tisíc let zpátky, dostaneme se do středověké Evropy. Z této doby nepřežila prakticky žádná politická instituce, snad kromě katolické církve, jejíž příklad naznačuje, že náboženské instituce mohou být trvalejší. Většina především lokální historie starší než tisíc let se ztrácí v mlze. Před invazí Normanů v roce 1066 je historie Anglie velmi kusá. Období starší než tři tisíce let je velkou neznámou, devět tisíc let překračuje dějiny současného lidstva."[16]

Cíl 7.1 Ochrana digitálních dat by měla být zajištěna po dobu, která překračuje existenci digitálního repozitáře

Příklady:

- Repozitář uzavírá smlouvu s jinou organizací, která je připravena převzít jeho data, pokud to bude třeba.
- Repozitář je v intenzivním kontaktu s jinými podobnými repozitáři.

Komentář:

Je asi dostatečně jasné, že zajistit uchování digitálních dat v použitelné podobě i poté, kdy už nebude existovat samotný repozitář, který je v současnosti skladuje, je velmi obtížné. Dnešní technologie neumožňuje uchovávat digitální data bez nutnosti nějaké aktivní údržby. Formáty musí být migrovány, úložná média je třeba vyměňovat, je třeba kontrolovat integritu dat atd. Pokud toto vše opomineme, naše data budou časem nepoužitelná.

V úvahu připadá jen jedna rozumná strategie, která spočívá v tom, že data ze zanikajícího repozitáře převezme a dále ochraňuje jiný fungující repozitář. Hlavním cílem plánu zajištění kontinuity je specifikovat smluvní ujednání týkající se převzetí dat v okamžiku, kdy původní repozitář zaniká. Ideálně by takový plán měl obsahovat:

- informace o repozitáři, který má data zdědit,
- specifikaci licenčních podmínek, za kterých bude materiál přijat do nového repozitáře,

16 BENFORD, Gregory. *Deep Time*. New York : Harper-Collins, 2000.

- specifikaci závazků nového repozitáře,
- formát dat a metadat, která je nový repozitář ochoten přijmout,
- specifikaci kompenzací, které nový repozitář za přijetí dat do svého systému získá.

V praxi může být obtížné vyjednat a udržovat takto podrobné podmínky nástupnické smlouvy. Proto je velmi důležité, aby repozitáře udržovaly těsné kontakty s jinými institucemi doma i v cizině, které by v případě potřeby mohly data převzít.

Repozitář (potažmo instituce) má sice určitý vliv na to, kdo data převezme, ale kontrola nad tím, co se děje s archivními objekty, končí v okamžiku předání dat do nového repozitáře (instituce). Kdo se bude o data starat, pokud i tento repozitář zanikne, je těžké odhadnout. Repozitář tak nemůže tušit, kdo se bude o jeho data starat o jednu nebo dvě generace později, ale může výrazně ovlivnit to, jak moc složité to bude. Repozitáře by si měly uvědomit, že jedním z hlavních důvodů, proč následovat mezinárodní standardy, je usnadnění práce repozitářů, které převezmou odpovědnost za uchovávaná data v budoucnosti.

5.8 Krizový plán

Plán pro případ krize (či katastrofy) má mezi plány strategických cílů zvláštní postavení. Ostatní plány strategických cílů se zabývají tím, jak zajistit bezvadný provoz repozitáře, plán pro případ katastrofy se zabývá situací, kdy je ohrožena samotná existence repozitáře.

Krizový plán by se měl zabývat i situací, kdy se repozitář zavře jednou provždy, což by ovšem mělo být předmětem plánu zajištění kontinuity. Neměli bychom přehlédnout, že je ještě další problém, kterým stojí za to se zabývat: jak zvládat ohrožení, ne pouze služeb nebo dat, ale repozitáře jako celku. Tato ohrožení nemusí být vždy přímo nepřátelská, mohou být také přirozeným důsledkem technologického vývoje nebo ekonomických změn.

Cíl 8.1 Digitální repozitář včas reaguje na podstatné změny prostředí

Příklady:

- Repozitář má vypracovanou konkrétní analýzu rizik a strategii managementu.
- Repozitář specifikuje procedury k zvládání předvídatelných katastrof.

Komentář:

Rizika prostředí, ve kterém repozitář existuje, mohou být zhruba klasifikována do těchto oblastí, nebo přinejmenším musí tyto oblasti brát v potaz:

- ekonomický otřes

Repozitáře, které mají jen několik málo základních zdrojů financování, zvláště ty, jež jsou součástí nějaké větší instituce, budou náchylné k finančním potížím. Obvyklý je nepravidelný příliv financí a může se ukázat, že je obtížné finance na další roky získávat.

Repozitáře, nebo přinejmenším vyspělé repozitáře, by měly mít vypracovány strategie zvládání výpadků financování. Obvyklými strategiemi jsou spolupráce s někým, kdo je ochoten poskytnout repozitáři po nějakou dobu finanční podporu, nebo vytváření finančních rezerv. Repozitář by také měl mít seznam služeb, jež poskytuje, seřazený podle důležitosti a měl by vědět, co je absolutně nezbytné zajistit. To by mělo usnadnit rozhodování o tom, jaké služby a činnosti zastavit v případě výpadku financování.

- politický otřes

Rozpad Sovětského svazu a východního bloku ukázal, že politický otřes se může týkat i zdánlivě stabilních společností. Repozitáře se v těchto situacích mohou ocitnout v rozporné situaci, kdy by měly

svoje sbírky uzavřít nebo naopak úplně otevřít. Politické otřesy také mohou vést k rabování nebo vyplenění repozitáře, jak ukazují svědectví z Bagdádských muzeí z nedávné doby.

- ztráta mandátu, důvodu existence

Mnohé repozitáře nejsou samostatné, jsou součástí větších institucí, ministerstev, firem nebo univerzit. Tyto organizace mohou být nuceny škrtat v rozpočtech nebo měnit svoji hlavní činnost, nové vedení (nebo vláda) nemusí považovat existenci repozitáře za důležitou. V takovém případě nemůže repozitář zůstat pasivní a musí se aktivně snažit těm, kdo mají rozhodovací pravomoc, dokázat svoji hodnotu a význam.

- technologický otřes

Technologické změny mohou velmi podstatně a rychle ovlivnit finanční plán i chování repozitáře. Např. vzestup užívání dnes již běžných MP3 přehrávačů mohl ovlivnit nároky na repozitáře specializující se na hudbu. To samé samozřejmě platí i pro technologické trendy hardwaru, softwaru a sítí. Repozitáře musí tyto trendy sledovat, posuzovat je a případně se adaptovat.

- přírodní katastrofa

Komunitu provozující repozitář mohou zasáhnout různé přírodní změny. Mnohé přírodní katastrofy se nedají předpovědět, ovšem jsou lokality, kde jsou různé výkyvy počasí obvyklé. Například v některých částech světa dochází k hurikánům a následným záplavám, zemětřesení jsou v některých oblastech častější. Instituce provozující repozitář by měla mít plány pro případ nejpravděpodobnějších přírodních katastrof a obvyklých katastrof jako jsou požáry, havárie topení nebo vody nebo pouhý výpadek dodávek elektrické energie. Je však nerealistické očekávat, že plány pokryjí všechny eventuální katastrofy. Základní vlastností katastrof je, že se objevují náhle, takže být připraven je důležité. Pro každý katastrofický scénář by měl plán říkat[17]:

1. jak zajistit bezpečnost osob přítomných v repozitáři,
2. jak maximalizovat řád, účinnost, rychlost reakce na katastrofu,
3. jak mobilizovat všechny zaměstnance, kteří mají nějakou roli v záchranných akcích,
4. jak minimalizovat ztráty dat a minimalizovat dobu, po kterou jsou služby repozitáře nedostupné uživatelům.

- ztráta uživatelů nebo příchod konkurence

Předpokladem existence každého repozitáře je, že existuje nějaká skupina uživatelů jeho služeb. Pokud uživatelé z nějakého důvodu zmizí, má to na repozitář zásadní dopad.

Repozitář poskytuje služby uživatelům. Je to vlastně určitý typ podnikání a samozřejmě zde také funguje konkurence. Dokonce i vládou zaštítěné repozitáře mohou být vytlačeny jinými digitálními repozitáři podporovanými například jinými ministerstvy. Příkladem mohou být repozitáře podporované ministerstvem kultury (má v kompetenci veřejné a národní knihovny) a těmi, které podporuje ministerstvo školství (má v kompetenci knihovny vzdělávacích institucí).

Repozitáře také mohou ztratit uživatele jednoduše proto, že se stanou nedůvěryhodnými. Může se stát, že uživatelé a vkladatelé dat nemusejí rozumět tomu, k čemu repozitář data používá. Zároveň služby, které repozitář poskytuje, mohou uživatelé začít považovat za nestabilní. Repozitář by proto měl mít jasnou komunikační strategii k jejich vysvětlení.

V každém případě by měl repozitář pravidelně vyhodnocovat uživatelské statistiky a všímat si služeb, které ztrácejí u uživatelů popularitu. Repozitáři také pomůže, budou-li uživatelé mít možnost ovlivnit nabízené služby nebo jejich design.

17 Tyto priority jsou převzaty z tohoto dokumentu:
Center of Southwest Studies. Disaster preparation and response plan [online]. Durango, Colorado (USA) : Fort Lewis College, 2007 - [cit. 2009-03-12]. Revised July 31, 2007. Dostupný z WWW: <http://swcenter.fortlewis.edu/Forms/DisasterPlan.htm>.

- ztráta klíčových zaměstnanců

Plán řízení lidských zdrojů se zabývá tím, jak předcházet ztrátě klíčových zaměstnanců, měli bychom nicméně být připraveni i na to, že tyto naše snahy selžou.

Ztráta zaměstnanců může mít různé dopady. Může dojít k odhalení obchodního tajemství repozitáře nebo může být ohrožena funkčnost služeb.

První případ je rizikem pro důvěryhodnost repozitáře jen tehdy, pokud jeho vnitřní fungování obsahuje nějaké nedůvěryhodné procesy, které dříve nebyly uživatelům známy. Pokud takové potenciálně nedůvěryhodné praktiky v repozitáři existují, je třeba mít připraveno tiskové prohlášení a další komunikační mechanismy pro případ, že dojde k jejich zveřejnění.

Druhý problém může být složitější. Cílem plánu řízení lidských zdrojů je této eventualitě předejít. Pokud přesto dojde ke ztrátě klíčového zaměstnance, je asi nejvhodnější dočasně ukončit poskytování některých služeb, dokud nebudou přijati a zaškoleni noví zaměstnanci. Dočasné přerušení některých služeb může mít negativní dopad na důvěryhodnost, ovšem poskytování nestabilních služeb ho bude mít rovněž. Někdy může být dočasné přerušení služeb uživatelům politicky nepřijatelné. V takových případech nezbývá nic jiného, než provoz omezit na prioritní služby repozitáře, zmíněné v kapitole o ekonomickém otřesu, a obětovat méně důležité služby a úkoly.

- bezpečnostní rizika

Ohrožena může být také samotná fyzická existence repozitáře. Ohrožení mohou být různého druhu v závislosti na typu dat, která repozitář ochraňuje, nebo na typu budovy, ve které sídlí. Repozitář sídlící v knihovně by mohl přijít o uživatele, kteří si myslí, že je budova ohrožena například teroristickým útokem. Repozitář ukládající citlivá data může být považován za nedůvěryhodný jen proto, že není ochráněn před fyzickou krádeží dat nebo hardwaru, na kterém jsou data uložena.

Repozitář by měl základní nebezpečí zmapovat a zařídit se tak, aby jim předcházel. Měl by mít také plány pro případ, že ke krádeži dojde.

5.9 Plán ochrany

Cíle plánů ochrany dat v repozitáři by měly být zaměřeny na období překračující bezprostřední krátkodobý, střednědobý nebo dlouhodobý management organizace. Implicitním cílem ochrany dat v repozitáři by mělo být zajištění dostupnosti a srozumitelnosti digitálních zdrojů po mnoho let, možná i „navždy“. Je proto obtížné formulovat obecné cíle ochrany dat tak, aby byly reálné a smysluplně měřitelné. Cílem většiny repozitářů bude samozřejmě uchovávat jeden nebo více druhů obsahu, pocházejícího z jednoho nebo více druhů zdrojů, a jedné nebo více skupinám uživatelů tento obsah zprostředkovávat. Úspěch lze vždy posuzovat jen pro určité časové období, což činí formulaci cílů dlouhodobé ochrany o to těžší. V krátkodobém horizontu se můžeme pokusit reálně odhadnout, kdy cílů ochrany dosaženo *nebylo*. Důkaz toho, že dlouhodobá ochrana byla úspěšná, je možné získat jen ve vzdálené budoucnosti, když budou data z repozitáře stále použitelná a dostupná. Pokud nemohou správci repozitáře garantovat, že jejich vlastní životy a infrastruktura ochrany dat, kterou vybudovali, budou trvat, dokud data neztratí hodnotu, je třeba dosažení cílů dlouhodobé ochrany hodnotit jinak.

Dlouhodobá ochrana digitálních informací je někdy popisována jako zajištění interoperability s budoucností. Možná stojí za to tuto myšlenku rozvinout. Interoperabilita je výzva i vzhledem k současnosti. Užitečným a měřitelným východiskem je požadavek, aby konkrétní informační objekt nebo třída informací byly uchovávány tak, aby byly použitelné na všech existujících platformách dostatečně širokou a různorodou skupinou uživatelů. Tento přístup je realistický, měřitelný a konkretizovatelný specifikací jednotlivých platforem (hardwarových i softwarových), na nichž mohou být informace uspokojivě zpřístupňovány a pochopeny, a explicitním vyjmenováním komunit (a jejich znalostních základen), které jsou schopny daným informacím porozumět. Repozitáře mají, jak jsme již zmínili, problémy prokázat spolehlivost a udržitelnost. Cíle dlouhodobé ochrany jsou naplněny, pokud jsou informace uchovány

spolu s kontextem a vysvětlujícími informacemi, jež usnadňují jejich použití uživatelskými komunitami a existujícími i vznikajícími technologickými platformami, navzdory nebezpečím a nepředvídatelným událostem.

Zavedení časového hlediska do plánování ochrany dat v repozitáři musí být promyšlené a realistické. Nemá smysl, aby repozitář plánoval, že bude udržovat svůj obsah navždy nebo dokud uložená data neztratí jakoukoli hodnotu. Lépe je cíle dlouhodobé ochrany rozdělit na kratší, pravidelné a předvídatelné časové úseky proložené hodnocením jednotlivých položek obsahu a rozhodováním o tom, které digitální objekty musí být nadále uchovávány. Převedením cílů dlouhodobé ochrany do řady krátkodobých nebo střednědobých cílů zdůrazníme současnou hodnotu informací, možnost přidat jim další hodnotu (což je hlavní heslo *digital curation*) a také aktivní, a nikoli pouze reaktivní, povahu dlouhodobé ochrany.

Cíl 9.1 Repozitář musí mít přehled o současném a vznikajícím hardwaru, softwaru a technologiích ukládání dat

Příklady:

- Udržovat a dokumentovat technické, sociální a právní analýzy operačních systémů Microsoft Windows XP, Linux, Mac OS X, Sun Solaris, Novell a dalších současných operačních systémů.
- Udržovat a dokumentovat technické, sociální a právní analýzy x86, AMD64, PowerPC, SPARC a dalších existujících modelů architektury procesorů.
- Udržovat a dokumentovat technické, sociální a právní analýzy optických disků, LTO pásek, optických pásek, SSD disků (solid state drives), hard disků a dalších existujících médií.
- Sledovat objevující se hardware, software a technologie vhodné k další analýze.

Komentář:

Aby mohli pracovníci repozitáře digitální data udržovat efektivně, musí mít dobrý přehled o současných a objevujících se trendech v oblasti technologií. Znalost by se neměla omezovat pouze na technické otázky, ačkoli ty jsou samozřejmě velmi důležité. Ovšem vedle sledování vývoje softwaru, hardwaru, architektury médií, výkonnosti, poruchovosti by se repozitáře měly zajímat i o právní otázky spojené s konkrétními nástroji a jejich používáním (tedy o autorská práva, často specifikovaná v uživatelské licenci) a také o další sociální otázky jako jsou rozšířenost používání, stabilita prodejců a o jakékoli identifikovatelné a obvykle používané kombinace technologií a modifikací.

Cíl 9.2 Repozitář by měl udržovat srozumitelné informace o všech strukturálních standardech (tj. například kódování souborů) a formátech

Příklady:

- Formálně dokumentovat technologické, sociální i právní charakteristiky každého přijatého nebo potenciálně přijatelného formátu souborů v plánu formátů.
- Formálně dokumentovat technologické, sociální i právní charakteristiky každého používaného nebo plánovaného archivačního formátu ve zvláštním plánu archivace formátů.

Komentář:

Analýza formátů je nezbytnou součástí procesu dlouhodobé ochrany, v žádném případě by ovšem neměla nahrazovat vlastní řešení digitální ochrany. Formáty jsou technickou strukturou, která dává informacím fyzickou podobu. K posouzení možností a nevýhod jednotlivých formátů je třeba zvážit řadu technických, právních a sociálních otázek. Ovšem strategie ochrany digitálních dokumentů založená pouze na analýze formátů je zcela nedostatečná. Podstatnou informační hodnotu má nejen struktura souboru, kterou se zabývá analýza formátů, ale velmi důležité jsou také sémantické vlastnosti každého informačního objektu, očekávání uživatelů a způsob, jakým data z repozitáře využívají.

Cíl 9.3 Repozitář musí udržovat znalosti o komunitě uživatelů, jejich kompetencích a znalostní bázi

Příklady:

- Formálně identifikovat a dokumentovat všechny části uživatelské komunity, mít přehled o dovednostech, znalostech a technologických nárocích uživatelů a jejich uživatelských očekáváních.
- Trvale udržovat přehled o uživatelské komunitě, udržovat dokumentaci, přidávat, odstraňovat a rozdělovat uživatelské komunity tam, kde je to nutné.

Komentář:

Jak již bylo řečeno výše, pouhé porozumění strukturálnímu kontextu každého informačního objektu nestačí k ochraně pro účely všech možných uživatelských komunit, které mohou chtít, aby jim byla informace srozumitelná. Je důležité, aby repozitář udržoval kontakty se všemi komunitami uživatelů, a je-li to vhodné, hledal nové potenciální uživatele. Je třeba mít neustále přehled o očekáváních a dovednostech uživatelů, přičemž každý aspekt ochranných aktivit by měl být ovlivněn především jejich potřebami. Analýza potřeb uživatelů může pomoci smysluplně definovat standardy strategie ochrany uložených dat. Repozitář musí mít na paměti, že bez ohledu na celkovou dobu ochrany dat je jeho cílem srozumitelně zprostředkovat ochraňovaná data svým uživatelům, přičemž množina uživatelů se v čase proměňuje, zvětšuje se a diverzifikuje.

Cíl 9.4 Repozitář by měl vědět, jaké nároky na dlouhodobou ochranu má každý typ uloženého informačního zdroje nebo třída uložených dat

Příklady:

Cíle je třeba přesně specifikovat:

Explicitně dokumentovat požadavky na obsah, chování, vzhled, kontext, srozumitelnost, interoperabilitu dat definováním minimálních a maximálních nároků přijatelné strategie ochrany dat.

- Explicitně formulovat požadavky na to, jak rozšířené mají být používané formáty, na ztrátovost/ /bezztrátovost, na toleranci chyb nebo na dokumentaci definováním minimálních a maximálních nároků přijatelné strategie ochrany dat.
- Explicitně formulovat požadavky na hardwarovou a softwarovou infrastrukturu a na zaměstnance definováním minimálních a maximálních standardů přijatelné strategie ochrany dat.
- Explicitně formulovat požadavky na procesy spojené například s tím, do jaké míry mají být strategie ochrany dat automatizované, ověřitelné nebo škálovatelné definováním minimálních a maximálních standardů přijatelné strategie ochrany dat.

Komentář:

Úspěch nebo neúspěch dlouhodobé ochrany dat může být smysluplně posouzen pouze, jsou-li požadavky na ochranu dat definovány přesně a promyšleně. Nástroje jako je například PLATO[18], plánování ochrany dat vytvořené v projektu PLANETS, mohou velmi usnadnit definování a třídění cílů dlouhodobé ochrany zaměřených na samotné ochraňované objekty, na kontext těchto objektů nebo procesy ochrany. Je důležité si uvědomit, že vhodnost strategie dlouhodobé ochrany je možné posoudit jen vzhledem ke specifickým vlastnostem jednotlivých digitálních objektů a požadavkům na jejich ochranu. Strategie založené pouze na technologii nebo analýze strukturálních vlastností digitálních objektů jsou nedostatečné: repozitář se například může rozhodnout kódovat všechny soubory Microsoft Word jako pouhý text. To je možné, pokud je hodnota obsahu všech objektů založena výhradně na jejich textovém obsahu. Zajímají-li se uživatelé například o formátování, stránkování, rozvržení, vnořené objekty nebo metadata souboru, mají smůlu. Tohle všechno nebude při použití dané zjednodušující strategie ochrany zachováno.

18 BECKER, Christoph. *The Planets Preservation Planning*...2008.

Cíl 9.5 Repozitář musí udržovat, provádět a hodnotit takové strategie dlouhodobé ochrany digitálních dat, které vyhovují konkrétním cílům dlouhodobé ochrany

Příklady:

- Vytvářet a hodnotit strategie dlouhodobé ochrany vzhledem k potřebám dlouhodobé ochrany.
- Pokud technologická, právní a sociální analýza či analýza uživatelské komunity odhalí zranitelnost existující infrastruktury dlouhodobé ochrany, je třeba ji dále rozvíjet a měnit.

Komentář:

Repozitář by měl mít dostatečné povědomí o vývoji kontextu své činnosti tak, aby mohl včas začít uskutečňovat nějakou ochrannou aktivitu. Rozmanitý vývoj může vyvolat potřebu určitých ochranných kroků. Technologie může přestat být k dispozici, prodejci mohou oznámit, že přestávají něco vyrábět nebo ukončují podporu prodaných produktů, vývoj legislativy může přinést další nebezpečí pro uchovávaný obsah, mohou se změnit očekávání, potřeby nebo schopnosti uživatelů. Repozitář by měl mít připravenu jednu nebo více strategií, jak ochránit srozumitelnost a hodnotu dat ve chvíli, kdy bude uchovávaný obsah ohrožen. Strategie dlouhodobé ochrany by měla být podrobována detailnímu hodnocení, ať už jde o emulaci, migraci nebo jakoukoli ze vznikajících strategií. Měla by být testována ve vhodném prostředí a výsledky by měly být porovnány s požadavky na ochranu konkrétních informačních objektů nebo tříd informací.

Cíl 9.6 Repozitář by měl na základě vypracované strategie pravidelně hodnotit, které informace mají být dále uchovávány

Příklady:

Vytvořit a aktualizovat kritéria umožňující rozhodovat, do jaké míry je konkrétní digitální obsah potřebné dále uchovávat a do jaké míry to odpovídá cílům repozitáře. To může zahrnovat například:

- nakolik odpovídá uchovávaný obsah smyslu existence organizace,
- nakolik uchovávání tohoto obsahu odpovídá strategiím organizace,
- je-li tento obsah autentický,
- je-li tento obsah použitelný,
- je-li tento obsah původní,
- stav a ucelenost tohoto obsahu,
- dostupnost kontextuálních informací k tomuto obsahu (například metadata),
- jeho přesnost.

Kritéria by se měla používat pravidelně a často. Na základě hodnocení je třeba odstraňovat a přesouvat obsah tak, aby byl soubor uchovávaných dat jako celek smysluplný, s ohledem na cíle a poslání repozitáře.

Komentář:

Neustálé hodnocení uchovávaných informací není přímo součástí procesu dlouhodobé ochrany dat. Přesto má význam definovat, jaké informační objekty nebo třídy informací by měly být dále archivovány. Zabudování tohoto odhadu do procesu dlouhodobé digitální ochrany funguje jako přirozené dělítko mezi jednotlivými ochrannými kroky, ověřující a ospravedlňující po sobě následující aktivity ochrany. Je to nástroj kontroly nad procesem dlouhodobé ochrany digitálních dat, zajišťuje trvalé povědomí o hodnotě digitálních zdrojů a o tom, nakolik vlastní proces dlouhodobé ochrany dat odpovídá širším cílům a prioritám organizace.

6 OD PLATTERU K DŮVĚŘE

PLATTER se zabývá výhradně managementem cílů a úkolů repozitáře. Sám o sobě nemůže zajistit důvěryhodnost, aniž by soupeřil s jinými projekty v této oblasti. Repozitáře by si měly uvědomit, že v současnosti neexistuje mezinárodně uznávaná autorita provádějící audit nebo certifikaci digitálních repozitářů. Jsou iniciativy, které se zaměřily na standardizaci auditu, třeba *Birds of Feather Group on Digital Repository Audit and Certification*[19]. Do té doby, než podobné skupiny vytvoří přijatelné standardy, je na jednotlivých repozitářích, aby definovaly, ve spolupráci se stakeholdery, své vlastní standardy a procedury směřující k dosažení důvěryhodnosti. To můžeme považovat za hlavní cíl repozitáře:

Cíl 0: Ve spolupráci se stakeholdery stanovit kritéria pro důvěryhodnost

Použitá kritéria by měla vycházet z existujících iniciativ v této oblasti, například manuály TRAC a Nestor nebo soubor nástrojů DRAMBORA, OAIS atd. Repozitář se například může rozhodnout, že si jako kritérium důvěryhodnosti vezme provedení sebehodnocení podle modelu DRAMBORA s tím, že výsledky tohoto hodnocení nechá dále zhodnotit dvěma nezávislými experty v oboru. Jiný repozitář může svoji důvěryhodnost založit na jednom ze zmíněných manuálů (TRAC, Nestor). Při volbě strategie auditu je třeba si uvědomit mimo jiné toto:

- Který z existujících nástrojů auditu je pro konkrétní repozitář nejvhodnější, je-li takový?
- Je externí audit nutný? A pokud ano, jaká kritéria použít k výběru externích auditorů?
- Kdy a jak často by se měl audit provádět?
- Je možné, aby výsledek auditu byl neúspěšný, a pokud ano, jaké plány je třeba udělat pro případ této eventuality?

PLATTER byl navržen jako podpora obou přístupů k auditu, jak dotazníkovému, tak sebehodnotícímu.

6.1 PLATTER a manuály TRAC + Nestor

Rozbor plánů strategických cílů v 5. kapitole tvoří podrobný seznam bodů, kterým se musíme věnovat při formulaci cílů a úkolů digitálního repozitáře. Jsou z velké části odvozeny z manuálů TRAC a Nestor. Neříkáme, která kritéria jsou z TRAC, která z Nestor a která z jiných zdrojů, takže repozitář, jenž si mezi nimi vybere své vlastní priority, může zjistit, že nevyhovuje všem kontrolním bodům ve vybraném manuálu. Repozitář, který bude používat TRAC nebo Nestor, by je měl používat od začátku plánování společně s PLATTERem. PLATTER se snaží být vyčerpávající v tom smyslu, že se snaží pokrýt všechny hlavní body identifikované v jiných nástrojích auditu. Repozitář, který přistoupí na filozofii PLATTERu a bude se věnovat všem bodům obsaženým v rozboru plánů strategických cílů v 5. kapitole, si může být jistý, že takto zahrnul do svého plánování všechny hlavní body z TRAC i Nestor.

6.2 PLATTER a DRAMBORA

PLATTER je myšlen jako doplněk nástroje DRAMBORA. Díky tomu by digitální repozitář naplánovaný za pomoci PLATTERu měl mít silnou výchozí pozici při využití sebehodnocení DRAMBORA. V sebehodnocení DRAMBORA repozitář musí identifikovat a zdokumentovat své cíle. Poté musí popsat jednotlivé kroky, které podniká k dosažení stanovených cílů, a zdroje, které k tomu využívá. Analýza rizik DRAMBORA pak pokračuje identifikací rizik, která ohrožují dosažení těchto cílů. Pokud byl k plánování cílů použit PLATTER, bude repozitář při analýze DRAMBORA v silnější pozici, protože jeho současné cíle budou podrobně zdokumentovány. Kombinace PLATTERu a DRAMBORA analýzy je mocným nástrojem k budování důvěryhodného repozitáře.

19 http://wiki.digitalrepositoryauditandcertification.org

Dodatek: Vazby mezi taxonomickými osami a plány strategických cílů

V této části se pokusíme zobrazit v tabulce některé základní vztahy mezi plány jednotlivých strategických cílů, které jsou popsány v 5. kapitole, a otázkami, které klasifikují typy repozitářů, jež jsou probrány v kapitole třetí. Obvykle každá jednotlivá taxonomická osa popsaná ve 3. kapitole souvisí s jedním nebo několika cíli repozitáře. Pouze otázku 4.2 týkající se standardů interoperability jsme nespojili s žádným konkrétním cílem. Nikoli proto, že by snad při formulaci cílů repozitáře nebylo třeba brát interoperabilitu v úvahu. Právě naopak, věříme, že interoperabilita bude časem stále důležitější ve všech oblastech fungování repozitářů. Rozhodli jsme se nevyjmenovávat všechny cíle spojené s interoperabilitou proto, že by vznikl široký seznam velmi se lišící pro jednotlivé repozitáře. Pokud se například repozitář rozhodne usilovat o interoperabilitu s jinými repozitáři v oblasti dlouhodobé ochrany (například v oblasti popisů formátů), bude to souviset s otázkou 4.2 plánu ochrany. Jiné repozitáře mohou využívat interoperabilitu při koordinaci vyhledávacích nástrojů, další například při automatizovaném získávání metadat. Každé z těchto použití interoperability bude vytvářet řadu vazeb uvnitř rámce PLATTERu. Předpokládáme tedy, že v budoucnosti, až bude pole interoperability jasnější, bude třeba otázku 4.2 přeformulovat tak, aby odpovídala vznikajícím rámcům interoperability.

Rovněž jsme se pokusili určit vzájemné vztahy mezi různými plány strategických cílů. Některé z nich mají tolik vazeb na jiné cíle, že vyjmenování všech vztahů by bylo matoucí. Příkladem může být cíl 3.1 Role zaměstnanců repozitáře. Protože předpokládáme, že všechny naše cíle jsou SMART, což znamená, že jejich naplňováním je možné někoho pověřit, pak musí být každý cíl spojen alespoň s jednou zaměstnaneckou rolí. Všechny nebo skoro všechny cíle jsou také spojeny s cílem 1.1, udržováním finančního plánu, protože každý cíl přináší nějaké náklady. Také jsme nevyznačili všechny vazby na datový plán, protože by jich bylo velmi mnoho. V zásadě jsou všechny cíle v datovém plánu spojeny přinejmenším s akvizičním plánem, plánem zpřístupňování a s plánem ochrany.

Cíl	Související cíle	Související otázky
1.1 Pravidelně sledovat a revidovat finanční plán	3.2 Získat a udržovat zaměstnance pro výkon specifických pozic	1.2 Komerční status
	4.1 Formulovat, udržovat a aktualizovat programové prohlášení	
	5.1 IT infrastruktura si musí umět poradit s rozsahem ukládání	
1.2 Udržet financování na takové úrovni, kterou vyžaduje běžný provoz repozitáře	1.4 Stanovit plány externí komunikace	1.1 Mandát
1.3 Vytvořit nouzové plány pro případ finančních omezení nebo krizí	7.1 Zajištění kontinuity	
	8.1 Reakce na změny prostředí	
1.4 Stanovit plány externí komunikace	2.1 Získávání relevantních dokumentů	1.1 Mandát
	4.2 Definovat komunitu/y uživatelů repozitáře	1.2 Komerční status
2.1 Získávání relevantních dokumentů	4.2 Definovat komunitu/y uživatelů repozitáře	1.1 Mandát
2.2 Vyjednat dohody o uložení	2.5 Zajišťovat aktuálnost smluv	1.3 Právní podmínky získávání obsahu
	1.1 Pravidelně sledovat a revidovat finanční plán	3.1 Metody akvizice

	6.1 Určit, jaké formáty digitálních objektů bude repozitář akceptovat (SIP)	4.1 Zdroje metadat
	6.1.2 Specifikovat zdroje a formáty bibliografických a deskriptivních metadat pro SIP	
	6.1.3 Specifikovat technická metadata pro SIP	
2.3 Získat fyzickou kontrolu nad dokumenty	5.3 IT musí garantovat dostupnost služeb	4.3 Strategie ukládání
2.4 Monitorování akvizice	6.1 Určit, jaké formáty digitálních objektů bude repozitář akceptovat (SIP)	3.2 Komplexnost dat
	6.2 Specifikovat formát dat a obsah metadat pro archivaci digitálních objektů (AIP)	3.3 Specializace dat
	6.4 Specifikovat transformaci SIP do AIP	4.4 Strategie softwarové podpory
2.5 Zajišťovat aktuálnost smluv		
3.1 Definovat zaměstnanecké pozice		2.3 Počet zaměstnanců
		1.2 Komerční status
		3.3 Specializace dat
		4.1 Zdroje metadat
		4.3 Strategie ukládání
		4.4 Strategie softwarové podpory
3.2 Získat a udržovat zaměstnance pro výkon specifických pozic		2.3 Počet zaměstnanců
		4.3 Strategie ukládání
		4.4 Strategie softwarové podpory
3.3 Rozvíjet kvalifikaci zaměstnanců		3.3 Specializace dat
		4.4 Strategie softwarové podpory
4.1 Formulovat, udržovat a aktualizovat programové prohlášení	1.1 Pravidelně sledovat a revidovat finanční plán	1.1 Mandát
	9.3 Repozitář musí udržovat znalosti o komunitě uživatelů	1.2 Komerční status
	1.4 Stanovit plány externí komunikace	3.5 Oprávnění k přístupu
4.2 Definovat komunitu/y uživatelů repozitáře	5.1 IT infrastruktura si musí umět poradit s rozsahem ukládání	1.1 Mandát
	5.3 IT musí garantovat dostupnost služeb	3.5 Oprávnění k přístupu
4.3 Formulovat a implementovat politiku zpřístupňování obsahu repozitáře	5.1 IT infrastruktura si musí umět poradit s rozsahem ukládání	1.2 Status
	6.2.1 Specifikovat metadata pro AIP	3.5 Oprávnění k přístupu

4.4 Specifikovat a realizovat technologické požadavky distribuce a zpřístupňování	6.3.1 Specifikovat metadata pro DIP	3.2 Komplexnost dat
	6.3 Specifikovat formáty dat pro digitální objekty distribuované uživatelům (DIP)	3.3 Specializace dat
	6.5 Specifikovat transformaci AIP do DIP	3.5 Oprávnění k přístupu
	4.2 Definovat komunitu/y uživatelů repozitáře	4.4 Softwarová strategie
	5.3 IT musí garantovat dostupnost služeb	
5.1 IT infrastruktura si musí umět poradit s rozsahem ukládání	1.1 Pravidelně sledovat a revidovat finanční plán	2.1 Množství dat
	3.3 Rozvíjet kvalifikaci zaměstnanců	2.2 Počet objektů
		2.4 Počet koncových uživatelů
		4.3 Strategie ukládání
5.2 Infrastruktura IT musí garantovat integritu a bezpečnost uložených dat	4.3 Formulovat a implementovat politiku zpřístupňování obsahu repozitáře	3.1 Metody akvizice
	3.3 Rozvíjet kvalifikaci zaměstnanců	3.4 Citlivost dat
	9.5 Repozitář musí udržovat, provádět a hodnotit takové strategie dlouhodobé ochrany digitálních dat…	3.5 Oprávnění k přístupu
		4.3 Strategie ukládání
		4.4 Strategie softwarové podpory
5.3 IT musí garantovat dostupnost služeb	9.1 Repozitář musí mít přehled o současném a vznikajícím hardwaru, softwaru a technologiích ukládání dat	3.2 Komplexnost dat
	3.3 Rozvíjet kvalifikaci zaměstnanců	3.3 Specializace dat
		3.5 Oprávnění k přístupu
		4.3 Strategie ukládání
		4.4 Strategie softwarové podpory
6.1 Určit, jaké formáty digitálních objektů bude repozitář akceptovat (SIP)	2.2 Vyjednat dohody o uložení	3.2 Komplexnost dat
		3.3 Specializace dat
6.1.2 Specifikovat zdroje a formáty bibliografických a deskriptivních metadat pro SIP	2.2 Vyjednat dohody o uložení	3.2 Komplexnost dat
		3.3 Specializace dat
		3.5 Oprávnění k přístupu
		4.1 Zdroje metadat
6.1.3 Specifikovat technická metadata pro SIP	2.2 Vyjednat dohody o uložení	3.2 Komplexnost dat
		3.3 Specializace dat

6.2 Specifikovat formát dat a obsah metadat pro archivaci digitálních objektů (AIP)	9.1 Repozitář musí mít přehled o současném a vznikajícím hardwaru, softwaru a technologiích ukládání dat	3.2 Komplexnost dat
	9.2 Repozitář by měl udržovat srozumitelné informace o všech strukturálních standardech	3.3 Specializace dat
	9.5 Repozitář musí udržovat, provádět a hodnotit takové strategie dlouhodobé ochrany digitálních dat…	
6.2.1 Specifikovat metadata pro AIP	9.5 Repozitář musí udržovat, provádět a hodnotit takové strategie dlouhodobé ochrany digitálních dat...	3.2 Komplexnost dat
		3.3 Specializace dat
		3.5 Oprávnění k přístupu
6.3 Specifikovat formáty dat pro digitální objekty distribuované uživatelům (DIP)	4.2 Definovat komunitu/y uživatelů repozitáře	3.2 Komplexnost dat
		3.3 Specializace dat
6.3.1 Specifikovat metadata pro DIP	4.2 Definovat komunitu/y uživatelů repozitáře	3.2 Komplexnost dat
		3.3 Specializace dat
		3.5 Oprávnění k přístupu
6.4 Specifikovat transformaci SIP do AIP		3.2 Komplexnost dat
		3.3 Specializace dat
		3.5 Oprávnění k přístupu
		4.4 Strategie softwarové podpory
6.5 Specifikovat transformaci AIP do DIP		3.2 Komplexnost dat
		3.3 Specializace dat
		3.5 Oprávnění k přístupu
		4.4 Strategie softwarové podpory
7.1 Zajištění kontinuity	1.3 Vytvořit nouzové plány pro případ finančních omezení nebo krizí	1.2 Status
		3.4 Citlivost dat
		3.5 Oprávnění k přístupu
8.1 Reakce na změny prostředí	5.2 Infrastruktura IT musí garantovat integritu a bezpečnost uložených dat	4.3 Strategie ukládání
	1.3 Vytvořit nouzové plány pro případ finančních omezení nebo krizí	
	9.5 Repozitář musí udržovat, provádět a hodnotit takové strategie dlouhodobé ochrany digitálních dat…	
9.1 Repozitář musí mít přehled o současném a vznikajícím hardwaru, softwaru a technologiích ukládání dat	3.3 Rozvíjet kvalifikaci zaměstnanců	4.3 Strategie ukládání
		4.4 Strategie softwarové podpory
9.2 Repozitář by měl udržovat srozumitelné informace o všech strukturálních standardech	6.2 Specifikovat formát dat a obsah metadat pro archivaci digitálních objektů (AIP)	3.2 Komplexnost dat
		3.3 Specializace dat
9.3 Repozitář musí udržovat znalosti o komunitě uživatelů	4.2 Definovat komunitu/y uživatelů repozitáře	3.2 Komplexnost dat
		3.3 Specializace dat
		3.4 Citlivost dat
		3.5 Oprávnění k přístupu

9.4 Repozitář by měl vědět, jaké nároky na dlouhodobou ochranu má každý typ uloženého informačního zdroje	3.3 Rozvíjet kvalifikaci zaměstnanců	3.2 Komplexnost dat
		3.3 Specializace dat
9.5 Repozitář musí udržovat, provádět a hodnotit takové strategie dlouhodobé ochrany digitálních dat...	6.2 Specifikovat formát dat a obsah metadat pro archivaci digitálních objektů (AIP)	3.2 Komplexnost dat
	6.2.1 Specifikovat metadata pro AIP	3.3 Specializace dat
	8.1 Reakce na změny prostředí	3.5 Oprávnění k přístupu
	5.2 Infrastruktura IT musí garantovat integritu a bezpečnost uložených dat	4.3 Strategie ukládání
		4.4 Strategie softwarové podpory
9.6 Repozitář by měl na základě vypracované strategie pravidelně hodnotit, které informace mají být dále uchovávány		1.1 Mandát

LITERATURA

A guide to the project management body of knowledge. Third Edition. Newtown Township (USA) : Project Management Institute, 2000.

BECKER, Christoph. *The Planets Preservation Planning workflow and the planning tool Plato* [online]. S.l. : PLANETS, 2008 [cit. 2009-03-12]. Dostupný z WWW: <http://www.planets-project.eu/docs/presentations/christoph_becker.pdf>.

BENFORD, Gregory. *Deep Time*. New York : Harper-Collins, 2000.

BURNER, Mike; KAHLE, Brewster. *Arc File Format* [online]. San Francisco (USA) : Internet Archive, 1996 [cit. 2009-03-12]. September 15, 1996, Version 1.0. Dostupný z WWW: <http://www.archive.org/web/researcher/ArcFileFormat.php>.

Center for Research Libraries (CRL), et. al. *Core Requirements for Digital Archives* [online]. Chicago (USA) : Center for Research Libraries, 2007 [cit. 2009-03-12]. Dostupný z WWW: <http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=92>.

Center for Research Libraries (CRL); OCLC Online Computer Library Center. *Trustworthy Repositories Audit & Certification : Criteria and Checklist* [online]. Chicago (USA) : Center for Research Libraries, 2007 [cit. 2009-03-12]. Version 1.0. February 2007. Dostupný z WWW: <http://www.crl.edu/PDF/trac.pdf>.

Center of Southwest Studies. *Disaster preparation and response plan* [online]. Durango (USA) : Fort Lewis College, 2007 [cit. 2009-03-12]. Revised July 31, 2007. Dostupný z WWW: <http://swcenter.fortlewis.edu/Forms/DisasterPlan.htm>.

Consultative Committee for Space Data Systems (CCSDS). *Reference Model for an Open Archival Information System (OAIS) : Blue Book, Issue 1.* Washington, DC : CCSDS Secretariat, 2002. Technická zpráva. CCSDS 650.0-B-1. January 2002. Recommendation for Space Data System Standards. Dostupný také z WWW: <http://public.ccsds.org/publications/archive/650x0b1.pdf/>.

Digital Curation Centre (DCC); DigitalPreservationEurope (DPE). *DRAMBORA interactive : Digital Repository Audit Method Based on Risk Assessment* [online]. Edinburgh : Digital Curation Centre, 2008. Dostupný z WWW: <http://www.repositoryaudit.eu/>.

GLADNEY, Henry M. *Preserving digital information*. Berlin : Springer-Verlag, 2007. ISBN 978-3-540-37886-0.

Nestor. *Catalogue of Criteria for Trusted Digital Repositories* [online]. Frankfurt am Main : Deutsche Nationalbibliothek, 2006 [cit. 2009-03-12]. December 2006. Version 1 (draft for public comment). Dostupný z WWW: <http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf>. URN:NBN:DE:0008-2006060703.

Open Archives Initiative. *The Open Archives Initiative Protocol for Metadata Harvesting* [online]. S.l.: Open Archives Initiative, 2002 [cit. 2009-03-12]. Protocol Version 2.0 of 2002-06-14. Document Version 2008-12-07T20:42:00Z. Dostupný z WWW: <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

VERHEUL, Ingeborg. *Networking for Digital Preservation : Current Practice in 15 National Libraries.* München : K.G. Saur, 2006. ISBN 978-3-598-21847-7.

Slovník zkratek

AIP	Archival Information Package (termín OAIS)
AMD64	označení 64bitových procesorů
ARC	archivní formát s bezztrátovou kompresí
CDDB	Compact Disc Database
CRL	The Center for Research Libraries
DCC	The Digital Curation Centre
DIP	Dissemination Information Package (termín OAIS)
DPE	DigitalPresevationEurope
DRAMBORA	Digital Repository Audit Method Based on Risk Assessment
DRM	Data Right Management – Správa práv k datům
DVD	Digital Video Disc
FP6	Šestý rámcový program EU
FTP	File Transfer Protocol
GB	Gigabyte
ID3 tag	formát metadat v hudebních souborech CD
IT	informační technologie
JHOVE	JSTOR/Harvard Object Validation Environment
JISC	Joint Information Systems Committee
JPEG	standardní formát a metoda ztrátové komprese používaný pro ukládání počítačových obrázků ve fotorealistické kvalitě
LIWA	Living Web Archives – výzkumný projekt
LTO pásek	Linear Tape Open – magnetopásková technologie
Mb	Megabit
MP3	formát ztrátové komprese zvukových souborů, založený na kompresním algoritmu MPEG
MPEG	Moving Picture Experts Group – formát videa
Nestor	The German Network of Expertise in digital long-term preservation
OAI–PMH	Open Archives Initiative Protocol for Metadata Harvesting – Protokol pro metadatovou interoperabilitu
OAIS	Open Archival Information System – Informační systém otevřených archivů
PDF/A	Portable Document Format – archivní verze formátu
PLANETS	Preservation and Long-term Access through Networked Services – výzkumný projekt
PLATO	Planets Preservation Planning tool – nástroj pro plánování dlouhodobé digitální ochrany vyvinutý v projektu PLANETS
PLATTER	Planning Tool for Trusted Electronic Repositoriem – Plán důvěryhodného digitálního repozitáře
RAID	Redundant Array of Independent Disks – vícenásobné diskové pole nezávislých disků
SHAMAN	Sustaining Heritage Access through Multivalent Archiving – výzkumný projekt
SIP	Submission Information Package (termín OAIS)
SSD disk	solid state drives
TIFF	Tag Image File Format – formáty pro ukládání rastrové počítačové grafiky
TRAC	Trustworthy Repositories Audit & Certification
WAV	Waveform audio format – formát zvukových souborů
x86	označení architektury procesorů
XML	Extensible Markup Language – rozšiřitelný značkovací jazyk

Obsah

Úvod k českému vydání PLATTERu	5
1. Shrnutí a úvod do PLATTERu	9
2. Důvěryhodný repozitář	10
3. Klasifikace repozitářů	12
4. Plánovací cyklus PLATTER	17
5. Plány strategických cílů PLATTERu	21
6. Od PLATTERu k důvěře	43
Literatura	49
Slovník zkratek	50

Průvodce plánem důvěryhodného digitálního repozitáře (PLATTER)
Repository Planning Checklist and Guidance

Colin Rosenthal, Asger Blekinge-Rasmussen, Jan Hutař, Andrew McHugh, Stephan Strodl, Emily Witham, Seamus Ross

Překlad: Jan Hutař, Ladislav Cubr, Marek Melichar
Odborná revize překladu: Andrea Fojtů
Grafická úprava: Martin Zhouf
Sazba: Jaroslava Černá
Tisk: KARTEX

Vydala Národní knihovna České republiky
1. vydání
Praha 2009