



# ***Proč jsou české digitální repozitáře „nespolehlivé“?***

Jan Hutař  
NK ČR

## Co je „důvěryhodnost“ ?



- prokázaná schopnost úložiště zachovat digitální dokumenty v dlouhodobém horizontu přístupné a použitelné, a to podle konkrétních kritérií
- pokud chce být úložiště navenek důvěryhodné, nejedná se pouze o technické řešení, právě naopak.
  - zajištění financování úložiště
  - kde a kdo úložiště provozuje
  - jak schopný se o něj stará personál
  - zabezpečení apod. (více viz níže)
- důvěryhodnost navenek – podstoupit nezávislý audit



# Certifikované důvěryhodné digitální úložiště



- data na něm budou uložena velmi dlouho > stále v použitelné formě
- Problém = překotný vývoj technologií (HW, SW) + zastarávání formátů (HW, SW)
- plánování ochrany digitálních dokumentů s ohledem na rychlý proces stárnutí a změn těchto dokumentů
- zajištění autenticity a zjištění stupně ohrožení
  - *klasické dokumenty – relativně snadné*
  - *digitální dokumenty - podstatně složitější*
- **ztráty v digitálním světě jsou rychlé, nevratné a ne vždy snadno a včas zachytitelné**

# Obecné problémy s dlouhodobou archivací digitálních objektů

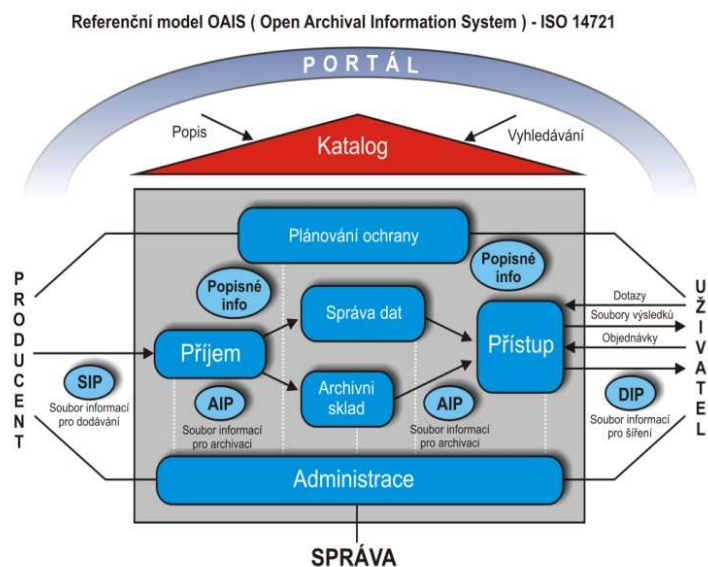


1. *Archivovaná informace musí být použitelná uživateli, kteří jsou vzdáleni v prostoru i čase a nemají podporu producenta té informace.*
  - producent informace už neexistuje – nepomůže
  - SW na kt. informace vznikla, už není podporován, není ani dokumentace
2. *Uživatelská komunita se bude během doby měnit.*
  - nová komunita nebude znát „pozadí“ vzniku informace > nevyužije ji
  - může pracovat v naprosto odlišném pracovním prostředí
3. *Archiv se bude měnit během doby.*
  - migrace na nové technologie – reorganizace, změny vztahů uvnitř
  - přenesení do jiné instituce - změny

# Základní funkce důvěryhodného digitálního úložiště



- v úplnosti je mapuje (a do detailů rozvíjí) referenční model OAIS
- úložiště je definováno jako **organizace**, která uchovává informace s cílem jejich zpřístupnění a využití
- *komplex funkcí podobný funkcím klasické knihovny, která **získává** (přijímá) dokumenty, **zajišťuje jejich popis, uložení a správu ve skladišti a jejich zpřístupnění** uživatelům.*



# Nástroje na certifikaci a audit důvěryhodných digitálních úložišť



- první kritéria hodnocení důvěryhodnosti digitálního úložiště:
  - *Trusted Digital Repositories : Attributes and Responsibilities* (RLG, OCLC, 2002)
- později další
  1. **DRAMBORA** (*Digital Repository Audit Method Based on Risk Assessment*)
  2. *NESTOR Criteria Catalogue*
  3. *Trustworthy Repositories Audit & Certification TRAC: Criteria and Checklist* (OCLC, CRL, 2007)
    - Organizace (řízení, struktura, udržitelnost, finance)
    - Správa digitálních objektů
    - Technologie, technologická infrastruktura, bezpečnost



# ***Kriteria hodnocení důvěryhodnosti digitálního úložiště***



Na základě výše uvedených nástrojů lze definovat 10 základních principů důvěryhodnosti:

Důvěryhodné úložiště ...

1. se musí zavázat k neustálému opatrování/správě digitálních objektů pro určitou cílovou komunitu.
2. musí prokázat svou životaschopnost (financování, personální otázky, struktury, procesů).
3. si musí osvojit a dodržovat potřebná smluvní a zákonná práva a dostát všem z nich plynoucím závazkům.
4. musí mít efektivní a dostačující rámcovou strategii.
5. získává a ukládá digitální objekty na základě stanovených kritérií, které odpovídají cílům a schopnostem instituce.

## ***Kriteria hodnocení důvěryhodnosti digitálního úložiště – pokr.***



...

6. neustále udržuje/zajišťuje integritu, autenticitu a využitelnost digitálních objektů, které trvale uchovává.
7. vytváří a uchovává potřebná metadata o událostech souvisejících s uloženými digitálními objekty v průběhu jejich uchování, jakož i metadata o samotném vytvoření digitálních objektů, podmínkách zpřístupnění a kontextu využití digitálních objektů.
8. musí naplnit nezbytné požadavky na zpřístupnění objektů ven z úložiště určité komunitě.
9. musí mít strategii pro plánování ochrany a souvisejících procesů včetně záchranných prací.
10. musí mít technickou infrastrukturu adekvátní pro účel neustálé údržby a zajištění digitálních objektů.

# ***Self audit a nástroj DRAMBORA***

***Digital Repository Audit Method Based on Risk Assessment***



- vznik počátkem r. 2007 – Digital Curation Centre (DCC) a DPE
- styčné body s certifikací TRAC i NESTOR
- DRAMBORA = nástroj v netradiční podobě
  - tvoří ho totiž v podstatě velmi podrobný návod a šablony k použití - v.1

nástroj pro *INTERNÍ* audit, který:

- pomůže instituci, kt. plánuje *EXTERNÍ* certifikaci svého úložiště
  - následný externí audit může pak být levnější, rychlejší, úspěšnější
- pomůže odhadnout možnosti úložiště, identifikovat slabé i silné stránky

[www.repositoryaudit.eu](http://www.repositoryaudit.eu)





## *Struktura celého procesu auditu*

DRAMBORA má 6 fází, které obsahují dohromady 10 úkolů

- 1. identifikace východisek (kontext organizace/instituce) - 2
- 2. identifikace strategie a regulačního rámce - 4
- 3. identifikace aktivit a prostředků - 1
- 4. identifikace hrozeb souvisejících s aktivitami a prostředky - 1
- 5. vyhodnocení hrozeb - 1
- 6. zvládnutí hrozeb - 1

možná rizika a hrozby jsou vyjádřeny pomocí číselných hodnot, což je ideální pro pozdější porovnání např. s dalším auditem



## *Okruhy a fáze auditu*

- Každý z úkolů je vypracováván z hlediska **8 okruhů** funkčnosti úložiště. Ve výsledku tedy dostáváme rizika, která jsou rozdělena podle těchto okruhů.

### **Procesní okruhy:**

- akvizice a ingest
- ochrana a uložení
- management metadat
- přístup a šíření

### **Podpůrné okruhy:**

- organizace a management
- personální otázky
- finanční management
- technická infrastruktura a bezpečnost



## ***Fáze auditu a úkoly 1***

### **Fáze 1 Identifikace východisek**

1. specifikujte mandát/poslání vašeho úložiště nebo organizace, kde je úložiště provozováno
2. vyjmenujte cíle a účel úložiště/organizace

### **Fáze 2 Identifikace strategie a regulačního rámce**

3. vyjmenujte dokumenty, které se týkají strategického plánování vašeho úložiště
4. uveďte všechny právní, smluvní a regulační rámce nebo dohody, které se buď i jen okrajově dotýkají vašeho úložiště
5. vyjmenujte dobrovolná nařízení, standardy a manuály, ke kterým se úložiště hlásí a řídí se podle nich
6. uveďte ostatní dokumenty a principy, se kterými je úložiště „ve shodě“

## ***Fáze auditu a úkoly 2***



### **Fáze 3 Identifikace aktivit a prostředků**

7. vyjmenujte všechny aktivity, prostředky potřebné k jejich provedení a jejich „majitele“ – tedy osoby odpovědné

### **Fáze 4 Identifikace hrozeb souvisejících s aktivitami a prostředky**

8. identifikujte hrozby spojené s aktivitami a prostředky vašeho úložiště

### **Fáze 5 Vyhodnocení hrozeb**

9. vyhodnoťte/ odhadněte hrozby – pomocí číselných hodnot z tabulky

### **Fáze 6 Zvládnutí hrozeb**

10. postavit se objeveným hrozbám, seznam opatření



## DRAMBORA audit v NK



- před vydáním nástroje DRAMBORA proběhly 2 zkušební audity v Evropě a 2 v USA.
- pilotní audity partnerů DPE – i NK – léto 2007
- v prvních 3 fázích (7 úkolů) jde o to shromáždit dostatek údajů k *identifikaci jednotlivých risků* (fáze 4) >> vyústí v *bodové ohodnocení jednotlivých risků* (fáze 5). V poslední fázi 6 by mělo dojít k *navrnutí určitého řešení*.
- všech 6 fází je zpracováváno vzhledem k již zmíněným 8 okruhům

## ***Procesní okruh Acquisition & Ingest***



- nejpalčivější problémy souvisí s tím, že úložiště NK nemá zatím nasazen žádný DOMS (Digital object management system).
- úložiště v podstatě zatím funguje jako „skladový“ prostor pro digitální dokumenty s file systémem.
- DOMS, podporující OAIS dodá úložišti potřebné vlastnosti, které se očekávají a mají pomoci zaručit dlouhodobou využitelnost uložených dokumentů a jejich formátů.
- takže stávající systém nepodporuje funkce, které by systém odpovídající OAIS podporovat měl, např. není schopen transformovat data, která do úložiště během Ingestu přicházejí, na AIP apod.

## ***Procesní okruh Preservation & Storage***



- opět DOMS > tj. nemáme archivní balíček AIP
- problémem je také otázka autenticity a integrity digitálních objektů v úložišti (neoprávněná chtěná i nechtěná manipulace s objekty apod.).
- další spornou otázkou je již řešená, ale nevyřešená problematika globálních persistentních identifikátorů, spolu s otázkou identifikátorů interních, která je řešena pouze částečně.
- dosti velká je i pravděpodobnost, že v budoucnu nebude možné zcela zavést požadovaný plán ochrany dokumentů, a to z různých důvodů (viz níže).

## ***Procesní okruh Metadata management***



- Problematika metadat náležejících k digitálním objektům představuje také mnoho rizik a hrozeb.
- digitální objekty nyní nemají ochranná metadata, s výjimkou Krameria.
- na vstupu do úložiště nějaká ochranná metadata jsou, nejsou ale pak doplňována (DOMS – po celou dobu existence objektu)
- zatím nedokončená definice jednotlivých balíčků SIP, AIP a DIP.
- nebezpečím je i to, že digitální objekty nebudou v budoucnu použitelné i přes všechna opatření (migrace, emulace apod.).

## ***Procesní okruh Access & Dissemination***



- v posledním procesním okruhu není celkové bodové ohodnocení hrozeb tak vysoké
- nejvíce bodů dostala nemožnost proměny AIP do balíčku určeného uživatelům (DOMS). Může ovšem být problémem kdykoliv.
- problémy mohou nastat i s migrací dat určených ke zpřístupnění
- copyright

# ***Podpůrný okruh Organisation & Management***



- celková koncepce NDK není dosud schválena vládou. To zásadně narušuje důvěru v úložiště, nejsou zajištěny organizační formality.
- zařazení a právní ukotvení úložiště ve struktuře NK
- problematika „digital preservation“ řešena v rámci několika oddělení
- chybí psaná a schválená dokumentace k procesům, které se okolo úložiště odehrávají
  - key staff member a žádná dokumentace > problém
- hrozbou může být i při rutinním běhu úložiště nekompetentnost vedoucích pracovníků, které může vést až ke ztrátě dat.



## *Podpůrný okruh Staffing*



- nedostatek kvalitního IT personálu
  - dostatečné financování celého úložiště a věcí okolo něj
  - schopnost a ochota managementu udržet v instituci schopné lidi
- ztráta klíčového zaměstnance (př. pouze on zná přístupová hesla, pouze on programoval celý systém, pouze on má odpovídající znalosti určité věci apod.).
- zaměstnanci si musejí doplňovat znalosti, protože pokrok jde velmi rychle kupředu.
- zaměstnance drží u konkrétní firmy (instituce) nejen peníze, ale i celková atmosféra a spousta dalších věcí ;-)

## ***Podpůrný okruh Financial management***



- z hlediska kvality úložiště a aktivního přístupu k digital preservation, je primární získání a implementace nějakého DOMS
- díky neschválené koncepci NDK je z finančního hlediska zakoupení DOMS nejisté
- závislost úložiště na odpovídajícím, pravidelném řádně ukotveném systémovém financování
- financování i po zprovoznění úložiště (nejen zakoupení HW a SW)
- negativní vliv jakéhokoliv snížení/ omezení financí na chod celého úložiště a jeho cíl

## ***Podpůrný okruh Technical infrastructure & Security***



- monitorování + funkčnost i zastarávání HW/SW - souvisí i se schopností pracovníků problém poznat a jednat
- možný nedostatek místa na úložišti (např. na zálohování).
- riziko poškození důsledkem nějaké nehody či neštěstí (bezpečnostní plán)
- omezení a monitorování fyzických vstupů na pracoviště
- Výpadek dodávek elektřiny nebo připojení k síti.



## *Tak kde je vlastně problém?*

- 1. organizační stránka instituce**
  - palčivý problém - ovlivňuje funkčnost a existenci celého úložiště
- 2. finanční zázemí**
- 3. neschválení konceptu NDK vládou**
  - problém důvěryhodnosti - projekt nemá pevné ani jisté financování!
- 4. neexistence DOMS**
- 5. metadata**
  - malá míra zastoupení tzv. ochranných metadat (vstup i uvnitř)
- 6. nedostatek IT odborného personálu**
  - finančně náročné - hledat cesty, jak tyto lidi zaangažovat
- 7. podpora vedení instituce/knihovny**
  - musí chápat důležitost a náročnost celé problematiky a podporovat ji
- 8. podpora státu v této oblasti – naprosto katastrofální**



## *Tabulka risků - ukázka*

Risk ID	Risk name	
R34	Overall concept of the National Digital Library is not submitted	21
R08	Our system doesn't provide transformation of submitted objects to archival packages	20
R10	Integrity and authenticity of the digital objects in the repository is not controlled.	20
R18	Unclear what is within AIP	20
R19	Identifier of digital objects is not persistent	20
R21	Preservation metadata for archived content are not acquired	20
R26	Documented change history is incomplete or incorrect	20
R43	Allocated insufficient resources for the activity	20
R52	Uncertainty of getting money for purchasing the DOM system for the repository	20
R47	Lack of sufficient number of appropriately qualified staff	20

<b>T8:</b>	<b>Identify risks associated with activities and assets of your repository</b>	
<b>T9:</b>	<b>Assess the identified risks</b>	
<b>T10:</b>	<b>Manage risks</b>	
<b>Risk Identifier:</b>	R34	
<b>Risk Name:</b>	Overall concept of the National Digital Library is not submitted	
<b>Risk Description:</b>	National Digital Library concept establish proposed NDL and includes funding from government.	
<b>Example Risk Manifestation(s):</b>	In case concept wont be submitted, the whole activity about the repository of the Czech National Library will be highly endangered and probably will never have trusted status.	
<b>Date of risk identification:</b>	1.7.2007	
<b>Nature of Risk:</b>	Physical environment	x
	Personnel, management and administration procedures	x
	Operations and service delivery	x
	Hardware, software or communications equipment and facilities	x
<b>Owner:</b>	Library Top Management	
<b>Escalation Owner:</b>	Library Top Management	
<b>Stakeholders:</b>	all	
<b>Risk Relationships:</b>	all	
<b>T9+T10 -Risk Probability:</b>	3	
<b>T9+T10 -Risk Potential Impact:</b>	7	
<b>T9+T10 -Risk Severity:</b>	21	



Děkujeme za pozornost

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