

**|** [www.ndk.cz](http://www.ndk.cz)

# **Dlouhodobá ochrana digitálních dokumentů a projekt NDK**

**Jan Hutař**

**Andrea Fojtů**

**Marek Melichar**

**Knihovny současnosti 14.9.2011 České Budějovice**

# Obsah prezentace

- problematika dlouhodobé ochrany digitálních dat (DD)
- LTP v NK a v projektu NDK
- nové standardy metadat

- základní ochrana dat nyní i v budoucnu (**ochrana bit-streamu**)
- ochrana použitelnosti dat v budoucnu (**ochrana logická**)
- **logická ochrana = procesy zajišťující, aby digitální objekty zůstaly v budoucnosti:**
  - vyhledatelné
  - přístupné (zobrazitelné)
  - využitelné znovu a znovu
  - pochopitelné (obsah a smysl)
  - autentické

- archivní dokument není navždy neměnná entita uložená v archivu
- cílem dlouhodobé ochrany není jen zachování bitů
- ale **HLAVNĚ** zajištění **použitelnosti** digit. dokumentů
  - to žádná sebelepší záloha nezařídí
  - to je podstata dlouhodobé ochrany digit. dat
- **záloha není dlouhodobá ochrana** – zachování bit-streamu je **pouze** nutný **předpoklad** pro uchování obsahu a použitelnosti dokumentů v budoucnu

- tj. dokumenty v archivu (LTP) musí být stále živé
  - reagovat na změny technologií, komunity, způsobů zpřístupnění
  - doplňování metadat, reprezentací dokumentů
  - migrace
  - změny logické struktury
  - zápis o všem do metadat
- je potřeba
  - mít digitální repozitář se SW aplikací - s moduly a funkcionalitou odpovídající OAIS – LTP systém
  - produkovat odpovídající metadata

## OAIS (ISO 14721:2003 – Open Archival Information System)

... s jednotlivými digitálními objekty musí být uchován nejen informační obsah uchovávaných objektů, ale také další informace o původu a historii změn dokumentu, o jeho kontextu a zdrojích potřebných k porozumění ...

- výraz „digital preservation“ již v 90.tých letech 20. století
  - spíše ochrana fyzických předloh pomocí digitalizace
  - na přelomu 2000 posun – ochrana samotných dig. objektů
- první aktivity v Austrálii (PADI projekt a web), USA, NL, UK
- reálná snaha o vytvoření LTP systému
  - od 2002 NK Nizozemí – spolupráce s IBM, systém e-Depot
  - Německá NK – využití LTP systému DIAS (IBM)
  - od 2006 NK Nového Zélandu – systém KRONOS > Rosetta (ExLibris)
  - 2008 dosud - velký rozvoj – open source nástroje/služby, komerční LTP systémy, nově i open source LTP systémy

- národní projekty – NDIIPP (USA) od 2000, PADI (Austrálie), NESTOR (Německo)
- organizace - DCC, DPC, JISC aj.
- konference – Archiving, iPRES
- zájem EU na řešení logické dlouhodobé ochrany DD
  - projekty FP5 (2003) Erpanet; FP6 (2006) - Planets, **DPE**, Caspar, Keep, Shaman, Scape aj.
- **LTP systém dnes řeší většina vyspělých knihoven a archivů světa (UK, DE, USA, NL, FIN, NOR, AUS, FR, PL, HU, SK, EST aj.)**



# Poučení [pro NK/NDK]

- NK ČR od roku 2006 součástí DP komunity
- komunita získala spoustu zkušeností > na nich je třeba stavět
- **vývoj okolo LTP spěje k vytváření politik, nástrojů, znalostí**
- **jedinou cestou k úspěchu na poli LTP je znalosti, služby a nástroje sdílet!**
- poučení z vývoje 1. generace LTP (NK NL)
  - základním požadavkem na LTP je otevřenost systému (integrace externích nástrojů a služeb)
  - flexibilita datového modelu a nastavení workflow
  - veřejná dokumentace – funkčnost, specifikace metadat

- konečně stabilizovaná ochrana bit-streamu
- archiv dat na filesystému
- HSM + několikeré zálohy (2x páska na 2 lokacích)
- žádný systém pro správu dat/metadat
- archivní balíček = data + metadata (DTD, MASTER+, nověji i technická a administrativní)
  - dokumenty jsou od okamžiku uložení „zamrzlé“ na úložišti
  - v případě úpravy metadat nebo struktury dat > úprava jen v aplikaci zpřístupnění

# Současná situace v NK 2.

- logická dlouhodobá ochrana DD jako taková v NK neexistuje
- nemáme systém pro správu dokumentů, tj. ani pro dlouhodobou ochranu
- minimální možnosti správy obsahu, analýzy obsahu úložiště, kontroly integrity
  - žádné filtrování, hledání apod.
- snaha o implementaci procesů ochrany do NK
  - metadata, kontrolní součty, DP strategie, PID, využívání externích služeb > NDK

- dlouhodobá ochrana dig. dokumentů (logická i bit-streamu) je jeden ze tří základních cílů NDK
- NK a MZK by měly získat odpovídající řešení pro správu archivních dig. dokumentů a jejich logickou dlouhodobou ochranu
- řešení musí odpovídat světovým trendům a standardům

# Nároky na LTP systém NDK

- musí odpovídat referenčnímu rámci OAIS
- schopen uložit jakákoliv data (archivní, nová; externí)
- uložení a práce s desítkami milionů digitálních objektů
- funkční a někde běžící systém (knihovna, archiv) s podobným objemem dat
- okamžitě dostupný komerční systém, žádný vývoj ani „bastlení“ z různých částí jiných systémů
- musí jít o LTP systém – ne jiný typ systému (CMS)
- aktivita producenta systému na poli dlouhodobé ochrany

## Nároky na LTP systém NDK 2.

- systém s relevantní roadmapou
- systém s relevantní uživatelskou komunitou (v oblasti LTP), která určuje jeho další rozvoj
- modularita – škálovatelnost dle objemů i typů dat
- systém musí využívat standardy, postupy a služby běžně využívané v oblasti LTP i služby budoucí; v různých modulech
- rule-based workflow – nastavení procesů (vstup, správa, zpřístupnění)

# Co máme za sebou

- 2 kola RFI (2008-2009)
- návštěvy knihoven s nasazeným LTP systémem (2009-2010)
- studii proveditelnosti (leden 2010)
- testování komerčních systémů Rosetta 2.1 a SDB 4 (2010)
  - k dispozici dokumentace, školení správce i technika (SDK)
  - vytvořeny základní ingest aplikace pro testovací data (PSP-SIP)
  - cíl – vidět zblízka funkcionalitu, zjistit zda jsme v NK schopni takový systém provozovat
- zadávací dokumentace na LTP systém (únor 2011)
  - funkční a nefunkční požadavky

# Testování LTP systémů

- odpovídají OAIS – všechny moduly včetně Preservation planning
- oba v max. míře využívají open source nástrojů a služeb
- oba systémy umožňují efektivní práci s daty/metadaty v archivu
  - doplňování metadat (ingest, archiv; po celý životní cyklus)
  - doplňování nových reprezentací digitálních dokumentů
  - nastavitelné reporty
  - logické přeskupování dokumentů nebo jejich struktury
  - **automatické rozeznání formátů a jejich migrace případná**
  - **identifikaci formátových rizik**
  - **plánování ochrany (monitorování >risk>hodnocení riziků>test set>testy>vyhodnocení>ostrá migrace)**
  - atd. atd.



# Nové standardy

- metadata pro digitalizaci (i archivaci) v NDK
- METS (úroveň čísla a svazku)
  - MODS, DC
  - PREMIS Object, Event, Agent
  - MIX
  - ALTO XML
  - strukturální metadata METS
- metadata z nové digitalizace, archivu, externích zdrojů (Google, e-PV aj.)
- transformační modul
- digitalizace historických dokumentů zůstane netknutá – je mimo projekt NDK
- ovlivní i jiné projekty – VISK7, ANL+

# NDK (LTP) a jeho přínos

- systém na správu obsahu úložiště – pro NK nutnost
- předpoklad řešení logické ochrany DD
- změna pohledu na ochranu digitálních dokumentů v knihovní (nejen) komunitě v ČR
- standardní metadata
- dopad na celou NK
  - „nový vítr“ pro NK
  - ovlivní většinu úseků
  - změna stylu práce (nutno vše dokumentovat, vytváření strategických dokumentů)

## ▪ komerční

- Safety Deposit Box (fa Tessella UK)
- Rosetta (fa ExLibris, Izrael)

## ▪ open source

- Fedora a její nadstavby
- XENA (NA Austrálie) <http://xena.sourceforge.net/>
- RODA (Portugalsko, Uni of Minho) <http://tinyurl.com/3ynyzs6>
- CRIB (Portugalsko, předchůdce RODA)
- ARCHIVEMATICA <http://archivematica.org/> (Unesco + Kanada); tool pack
- MOPSEUS – založeno na Fedoře, Řecko
- HOPPLA – vývoj TUW Vídeň, pro malé instituce nebo domácnosti
- ePRINTS – Univerzita Southampton

## ▪ nástroje na preservation planning

- PLANETS testbed, PLATO aj.

**Děkuji za pozornost  
a ... držte palce ;-)**

**■ Statistice knih a periodik online!**

**[jan.hutar@nkp.cz](mailto:jan.hutar@nkp.cz)**