



# **Open source nástroje pro dlouhodobou ochranu digitálních dat**

Mgr. Andrea Fojtu

Odbor digitální ochrany

Národní knihovna ČR



# Open source nástroje

- **HOPPLA** (Rakousko, Technische Uni Wien)
- **MOPSEUS** (Řecko, Digital Curation Unit)
- **RODA** (Portugalsko, Portuguese National Archives) + **CRIB** (Portugalsko, Uni of Minho)
- **ARCHIVEMATICA** (Unesco + Kanada)
- **AIDA** (Velká Británie, University of London + JISC)

# HOPPLA (<http://www.ifs.tuwien.ac.at/dp/hoppla/>)



- **Home and Office Painless Persistent Long-term Archiving**
- vhodné řešení pro malé instituce a SOHO (**S**mall **O**ffice/**H**ome **O**ffice), není nutné sledovat nejnovější trendy v oblasti formátů a nástrojů pro dlouhodobou ochranu
- preferuje metodu migrací formátů
- funguje na podobném principu jako např. antivirový program
  - znalosti od expertů převedeny do balíčků „updatů“ pro jednotlivé instalace
- kombinuje:
  - **ochranu bit streamu** - základní ochrana dat nyní i v budoucnu
    - ochrana před fyzickou ztrátou dat (zálohování)
  - **logickou ochranu** - ochrana použitelnosti dat v budoucnu
    - logická ochrana = procesy zajišťující, aby digitální objekty i do budoucna **vyhledatelné, přístupné** (zobrazitelné), **využitelné** a **pochopitelné** (obsah a smysl)
      - pravidelným vyhodnocením



# HOPPLA #2



- problémy, ktoré sa musí dořešiť:
  - vyšší chybovosť (vývoj automatického mechanizmu pro řešení chyb), různorodost formátů v malých institucích (ve větších je tendence k standardizaci), migrační nástroj PS2PDF
- základní předpoklady:
  - Windows, Linux, Mac OS (Public Release 2.0), JRE 1.6



## POSTUP:

1. **profil sbírky** (collection profile) – technické charakteristiky (formáty objektů, detailnější popis) + agregace metadat
2. **požadavky uživatele** (user's requirements) – rating 0-4
  - 0 (neprovádět žádné ochranné akce) až 4 (sledovat a provádět všechny akce)
  - 2,3 nabízí migraci z proprietárních formátů do open source formátů
3. **systemový profil** (system profile) – kontrola dostupných migračních nástrojů na straně uživatele, nabídne k doplnění chybějící (portable migration)



# Mopseus (<http://194.177.192.14/mopseus/>)



- Digital Repository System with semantically enhanced preservation services
- digitální repozitář (základ Fedora bez nutnosti dalšího programování a customizace):
  - robustní datový model (namapovaný na datový model PREMIS); sémantická anotace interní struktury DO
  - nenabízí interní migrační nástroj pro migraci formátů (z důvodu omezených finančních prostředků na vývoj)
  - migrační aplikace pro DSpace repozitáře
- základní předpoklady:
  - Windows XP/Vista, JDK
- **plány do budoucna:**
  - průvodce pracovních postupů (workflow wizards) pro snadnější definici, plánování a provedení aktivit pro správu objektů
    - uživatelsky přívětivější rozhraní
  - API jako náhrada za interní migrační nástroj (napojení na externí služby typu PLATO)



# RODA (<http://roda.di.uminho.pt/#home>) & CRIB



- RODA – **Repositório de Objectos Digitais Autênticos**
  - SOA – servisně orientovaná architektura (“balíčkový systém” interoperabilních služeb, které je možné využít v různých systémech; prostředkem pro integraci rozdílných systémů)
  - cílem je poskytnout technické řešení na úrovni národního archivu
  - CRiB (Conversion and Recommendation of Digital Object Formats) migrační podpora integrovaná do RODA
- **respektuje standardy**
  - OAIS (DO jako binární obsah a vztahy a reprezentace), METS, EAD, PREMIS
    - Fedora Commons jako jádro systému
- RODA vs. Mopseus
  - podobný přístup (Fedora)
    - DataStreams v Mopseus nejsou součástí METS
    - Mopseus má jednodušší instalační postup (malé až středně velké instituce)
- **prezervační možnosti (CRiB)**
  - identifikace formátů
  - doporučení optimální verze pro migraci
  - konverze digitálních objektů do nových, vhodnějších formátů
  - kontrola a vyhodnocení migračního procesu
  - vygenerování ochranných metadat v PREMIS



# ARCHIVEMATICA (<http://archivemata.org/>)



- ve vývoji, beta verze pro implementaci do partnerských institucí začátkem roku 2011
  - UNESCO Memory of the World Subcommittee on Technology poskytlo externí financování
    - prozatím možnost „individuálního“ testování
- „micro-services“ vybudovaných kolem souborového systému
  - cílem je redukce technické komplexnosti a údržby
  - zjednodušení vývoje
  - možnost snadno nahradit zastaralé prvky systému
    - dohromady vyhovují OAIS modelu  
<http://archivemata.org/wiki/images/d/dc/Archivemata-architecture-7May2010-2.png>
    - RODA (kde opětovné využití prvků je na makro úrovni (služby), nikoliv na úrovni mikro (třídy) jako u Archivemata)
    - Virtual appliance – upravený Ubuntu Linux s open source nástroji
- **24 “micro-services”:**
  - **kategorie:** receive SIP, reviewSIP, quarantineSIP, appraiseSIP, prepareAIP, reviewAIP, storeAIP, provideDIP, monitorPreservation



# AIDA (<http://aida.jiscinvolve.org/wp/>)



- představuje metodologii pro hodnocení činností souvisejících s repositářem
  - kapacity, stav připravenosti, celková schopnost správy digitálních sbírek
  - **hodnocení, nikoliv audit!**
    - je to o evaluaci (measurement); „důkazy“ pro aktivity nejsou nutné; špatný výsledek tedy neexistuje
    - na rozdíl od např. DRAMBORA **kontrolu v instituci neprovádí jedna osoba/ auditor**
      - všechny zodpovědné osoby vyplňují hodnotící tabulky (na institucionální úrovni a úrovni oddělení – institutional vs. departmental level)
- produkty pro doporučující akce jsou na řadě
- **struktura :**
  - **prvky** (elements) – popisují jednotlivé aspekty správy digitálních sbírek (organizational, technical, resources leg)
  - **fáze** (stages) – stupeň připravenosti (1-5)
    - institucionální úroveň, úroveň oddělení
    - např.: Prvek č. 6 (Sustainability of Funding), Fáze 2. „institucionální úroveň: „Funding comes in as short-term, one-off grants and ad-hoc awards.“

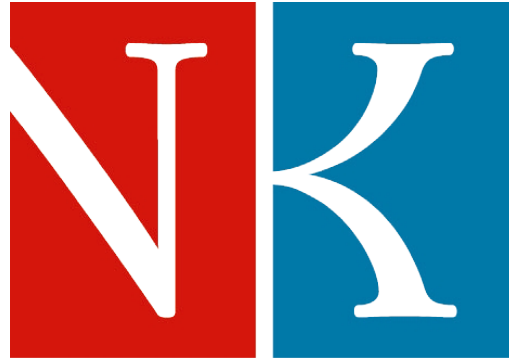




# Přehled nástrojů



- **Fedora a její nadstavby:**
  - RODA (Portugalsko, Portuguese National Archives) - <http://roda.di.uminho.pt/#home>
  - CRIB (Portugalsko, Uni of Minho) - <http://crib.dsi.uminho.pt/>
  - MOPSEUS (Řecko, Digital Curation Unit) – <http://194.177.192.14/mopseus/>
- **další nástroje:**
  - HOPPLA (Rakousko, Technische Uni Wien) - <http://www.ifs.tuwien.ac.at/dp/hoppla/>
  - ARCHIVEMATICA (Unesco + Kanada) - <http://archivemata.org/>
    - XENA (NA Austrálie) <http://xena.sourceforge.net/>
- **nástroje pro plánování dlouhodobé ochrany**
  - PLANETS testbed, PLATO apod.
  - AIDA (Velká Británie, University of London + JISC) - <http://aida.jiscinvolve.org/wp>



**Děkuji za pozornost  
Otázky?**

[andrea.fojtu@nkp.cz](mailto:andrea.fojtu@nkp.cz)  
[andrea.fojtu@mac.com](mailto:andrea.fojtu@mac.com)

<http://www.ndk.cz>