



komplexní řešení pro dlouhodobou archivaci  
digitálních (knihovných) sbírek

## ARCLib – LTP řešení pro knihovny

Prezentace je dostupná pod licencí Creative Commons, licence: [CC-BY-SA-4.0](https://creativecommons.org/licenses/by-sa/4.0/), via <http://repozitar.techlib.cz/record/1297>

Mgr. Eliška Pavlásková, Ph.D.

PhDr. Zdeněk Vašek, Ph.D.

Knihovna Akademie věd ČR

11. Konference o šedé literatuře a repozitářích

Praha, 24. 10. 2018



# Základní informace

- Výzkumný projekt financovaný z programu MK ČR NAKI
- 2016-2020
- 23 mil. Kč
- Spolupráce Knihovny AV ČR, Národní knihovny ČR, Moravské zemské knihovny v Brně a Masarykovy univerzity
- **Vývoj open source řešení pro dlouhodobou archivaci digitálních dokumentů (dat) ARCLib**

[www.arclib.cz](http://www.arclib.cz)



## Cíle projektu

- **Open source** řešení ARCLib pro dlouhodobou archivaci digitálních dat
- Metodika pro dlouhodobou **logickou ochranu** digitálních dat
- Metodika pro **bit-level ochranu** a návrh řešení pro fyzické ukládání digitálních dat
- **Ověření v praxi** formou poloprovozu

# Předpokládané vlastnosti řešení

- **Interoperabilita s NK ČR**
  - Podpora standardů digitalizace NDK
- Otevřené řešení předpokládající možnost **rozšíření pro další oblasti**
  - **Zapojení dalších typů dat** – SIP profil producenta
  - **Další vývoj systému** – přizpůsobení potřebám např. v prostředí archivů
- **Open source** alternativa ke komerčním systémům
- Vhodné primárně pro **národní, specializované a krajské knihovny**
  - Doplněno **metodikami** pro implementaci a provoz systému

# Metodika pro logickou ochranu dat

- Obecná/teoretická část
  - Popis problematiky dlouhodobého ukládání, popis základních standardů, které se v této oblasti používají a obvyklé strategie a postupy.
- Praktická část vázaná na produkt ARCLib
  - Postup praktické implementace požadavků na dlouhodobou archivaci při návrhu konkrétních SW řešení.
  - Popis systému s ohledem na implementaci.
- Implementační část
  - Doporučení pro uživatele systému ARCLib pro jednotlivé typy dat.
- Certifikována 2017 - předpoklad pravidelné aktualizace



# Metodika pro bitovou ochranu dat

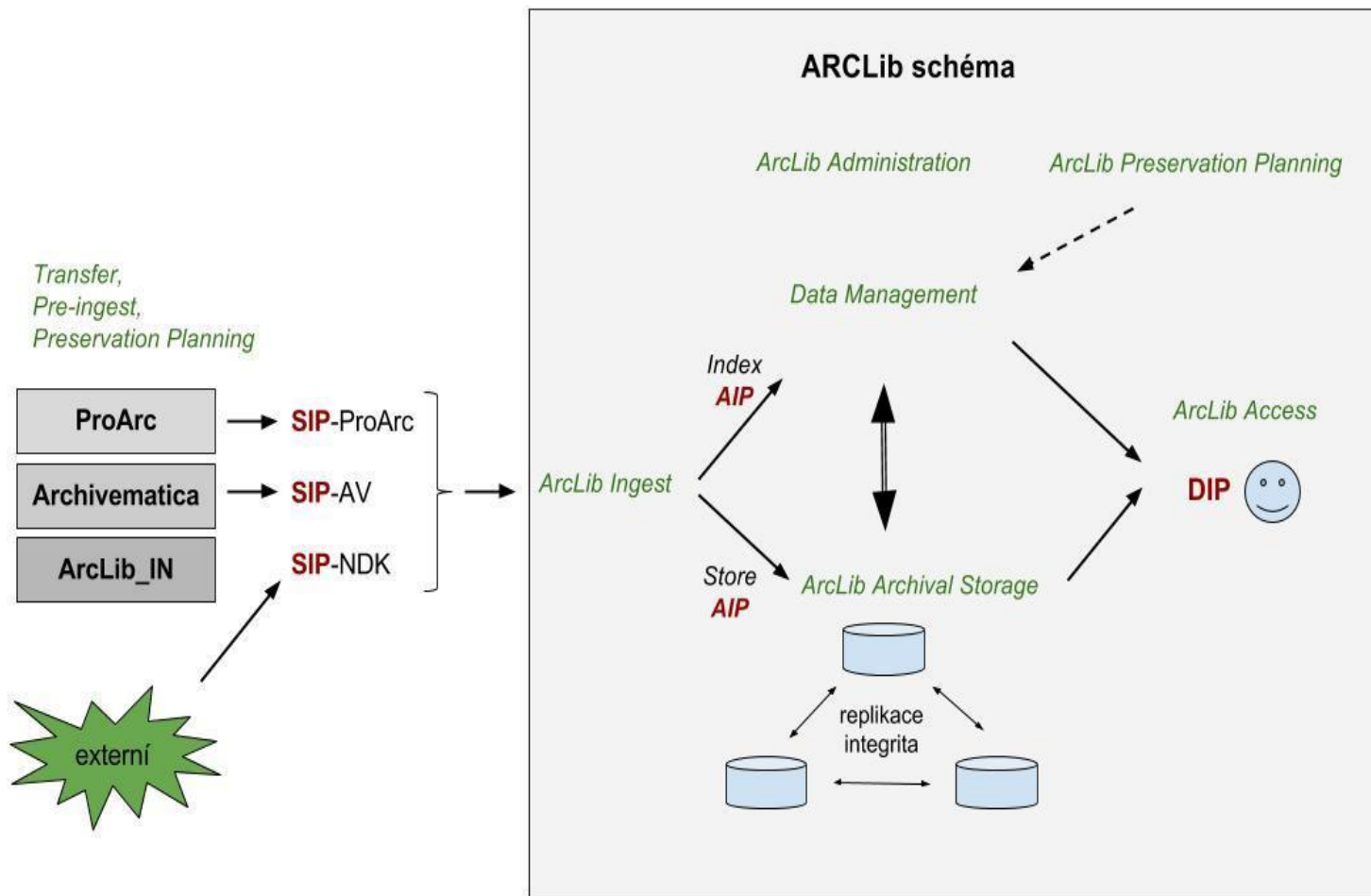
- Obecná/teoretická část
  - Strategie dlouhodobé ochrany
  - Zákonné požadavky na uchovávání digitálních archiválií
  - Shrnutí dobré praxe a požadavků na certifikaci úložiště
- Implementace strategií v systému ARCLib
- Zář 2018 – odeslána k certifikaci



# Popis zastřešující aplikace ARCLib

- Implementace funkčních modelů OAIS (ČSN ISO 14721)
- Neopakuje to, co už nástroje typu ProArc nebo Archivematica již umí (využití pro tvorbu SIP balíčků).
- Vývoj ARCLib se zaměřuje na vytvoření základních modulů:
  - ARCLib Ingest
  - ARCLib Data management
  - ARCLib Archival storage
  - ARCLib Administration
  - ARCLib Access

# Schéma







# ARCLib Ingest

- Vstupem plnohodnotný SIP
- Systém je schopen zpracovat data v následujících strukturách:
  - ProArc NDK monografie
  - ProArc NDK periodika
  - ProArc nativní monografie a periodika
  - ProArc zvukové dokumenty
  - NDK periodika a monografie
  - Archivematica DSpace
  - Archivematica Obecný
  - NDK elektronické dokumenty
- Funkce
  - Validace SIP
  - Extrakce metadat ze SIP
  - Tvorba metadat
  - Zpracování na základě profilu producenta
- ARCLib AIP XML
  - Metadatový profil založený na METS a PREMIS



# ARCLib Data management

- Správa AIP
- Vyhledávání a indexace (dostupné též přes API)
  - Popisná metadata
  - Technická metadata
  - Administrativní metadata
- Prohlížení obsahu AIP
- Editace metadat
- Reporting

- Users
- Producers
- Producer profiles
- Ingest
- Ingest routines
- Ingest batches
- Validation profiles
- Sip profiles
- Storage administration
- Workflow definitions
- Deletion requests
- Search queries
- AIP search**

## AIP search

## Sort

Producer ID

## Producer ID

Contains

## User ID

Contains

## State

Contains

## SIP version number

Greater or equal  Less or equal

## ID of previous SIP version

Contains

## XML version number

Greater or equal  Less or equal

## ID of previous XML version

Contains

## Document

Contains

## Root

## Label

Contains

## Type

Contains

## Sip ID

Contains

## Header

## Created

from  to

## XML ID

Contains

## Authorial ID

Contains

## Descriptive metadata

## Generic Dublin Core

Contains

## Specific sets of Dublin Core

Contains

## Aggregated extracted technical metadata

## Formats

## Date created by application

from  to

## File format

Contains

## Format registry key

Contains

## Format registry name

Contains

## Creating application name

Contains

## Creating application version

Contains

## Preservation level value

Contains

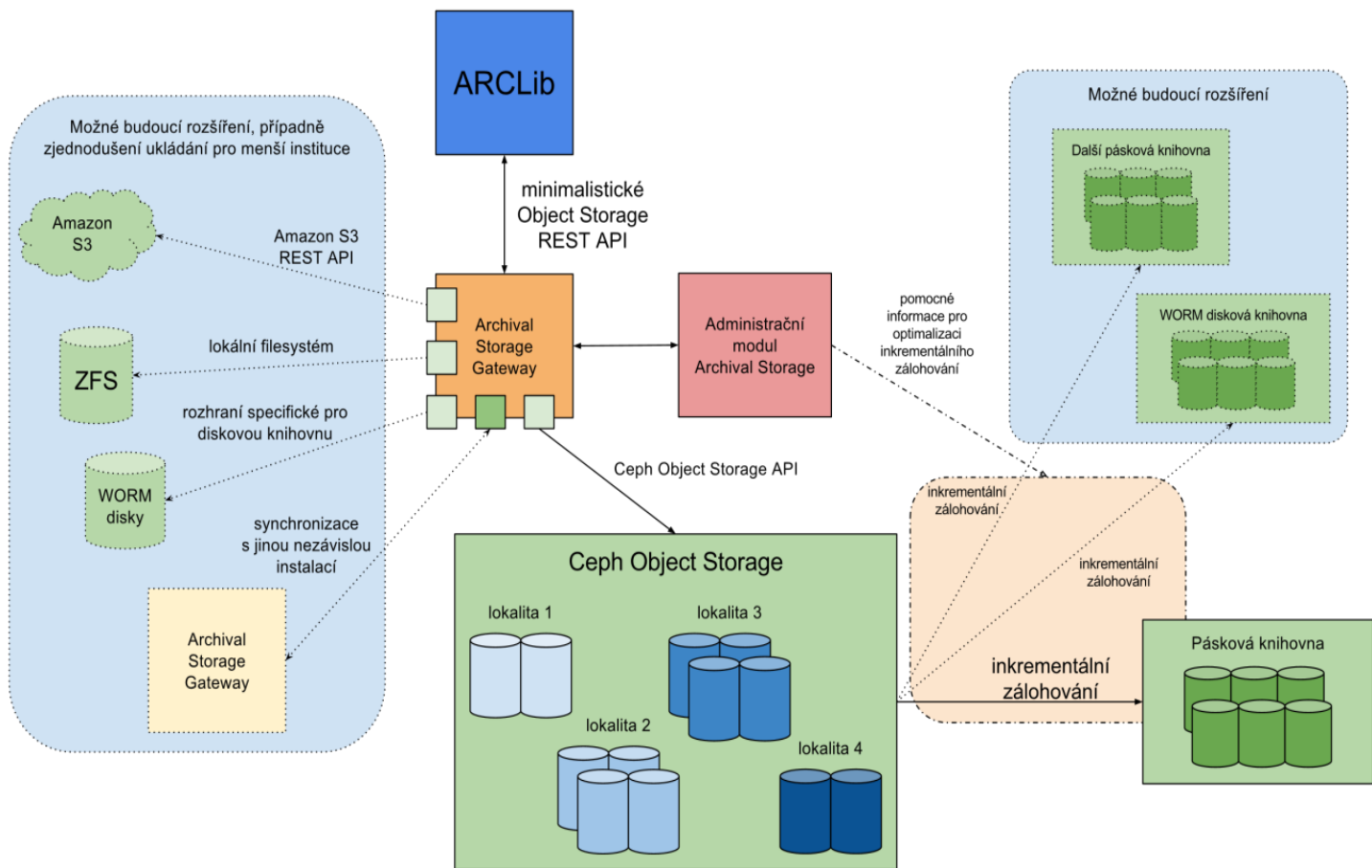
## Scanner model serial no

Greater or equal  Less or equal

## File count

Greater or equal  Less or equal

# ARCLib Archival storage



Users

Producers

Producer profiles

Ingest

Ingest routines

Ingest batches

Validation profiles

Sip profiles

Storage administration

Workflow definitions

Deletion requests

Search queries

AIP search

## Storage administration / local storage



Delete

Storage

[Synchronization information](#)**Name**

local storage

**Host**

localhost

**Port**

0

**Priority**

10

**Storage type**

FS

**Configuration file**

```
{  
  "rootDirPath": "/opt/archival-storage/data"  
}
```

**Poznámka** Write only Reachable

Storno

Save and close



# ARCLib Administration

- Konfigurace workflow zpracování na příjmu
  - Související registry – registr validačních profilů, registr využívaných scriptů...)
- Správa nástrojů třetích stran
- Správa uživatelů a nastavení autentikace



# ARCLib Access a Preservation Planning

- **Dark archive** - není učený pro koncové uživatele
- Přístup k datům omezen na možnosti exportu – **DIP je roven AIP**
- Registr formátů – napojení na **PRONOM**
- Další funkce plánování ochrany přesunuty vně systému (předpokládá se **metodické působení NK ČR**)



# Současný stav (říjen 2018) projektu

- Certifikována a zveřejněna **Metodika pro logickou ochranu dat** - <http://hdl.handle.net/11104/0282107>
- **Metodika pro bitovou ochranu dat** odeslána k certifikaci
- Zprovozněna interní **testovací verze ARCLib** na infrastruktuře Knihovny AV ČR
- **2020** – předpoklad finálního řešení pro dlouhodobou archivaci ARCLib a jeho ověření poloprovozem v Knihovně AV ČR





**Děkuji za pozornost**

Mgr. Eliška Pavlásková, Ph.D.  
[eliska.pavlaskova@ruk.cuni.cz](mailto:eliska.pavlaskova@ruk.cuni.cz)

PhDr. Zdeněk Vašek, Ph.D.  
[zdenek.vasek@ruk.cuni.cz](mailto:zdenek.vasek@ruk.cuni.cz)