



# Digitální knihovny a repozitáře

Daniel Jakubík  
<jakubik@fi.muni.cz>



Obecný úvod

Identifikátory

Metadata

Interoperabilita

- 1 Obecný úvod
- 2 Identifikátory
- 3 Metadata
- 4 Interoperabilita



# Co je to digitální knihovna?

Obecný úvod

Identifikátory

Metadata

Interoperabilita

Problém s **vymezením pojmu**:

- neustále se vyvíjí
- různé úhly pohledu
- nejednotné názvosloví



# Co je to digitální knihovna?

Obecný úvod

Identifikátory

Metadata

Interoperabilita

Problém s **vymezením pojmu**:

- neustále se vyvíjí
- různé úhly pohledu
- nejednotné názvosloví

Po masovém rozšíření webových technologií se objevili i názory, že celý Internet je vlastně digitální knihovnou.



# Co je to digitální knihovna?

Obecný úvod

Identifikátory

Metadata

Interoperabilita

Problém s **vymezením pojmu**:

- neustále se vyvíjí
- různé úhly pohledu
- nejednotné názvosloví

Po masovém rozšíření webových technologií se objevili i názory, že celý Internet je vlastně digitální knihovnou.



Obecný úvod

Identifikátory

Metadata

Interoperabilita

## Pohled pracovníka v IT

*Digitální knihovna je spravovaná sbírka informací spolu s odpovídajícími službami, přičemž informace jsou uloženy v digitální podobě a jsou dostupné prostřednictvím sítě.*

- W.Y.Arms, 2000



Obecný úvod

Identifikátory

Metadata

Interoperabilita

## Pohled knihovníka

*Digitální knihovny jsou organizace, které poskytují zdroje (včetně specializovaného personálu) umožňující provádět výběr, strukturování a zpřístupnění sbírek digitálních prací, tyto práce dále distribuovat, udržovat jejich integritu a dlouhodobě uchovávat – a to vše s ohledem na snadné a ekonomické využití určitou komunitou nebo množinou komunit uživatelů.*

*- Digital Library Federation, 1997*



# Srovnání digitálních a klasických knihoven

Obecný úvod

Identifikátory

Metadata

Interoperabilita

- dlouhodobě budovaná a spravovaná sbírka dokumentů vybraných na základě stanovených kritérií
- tvorba metadat používaných pro sestavování katalogů, rejstříků
- poskytování služeb určité komunitě uživatelů





# Srovnání digitálních a klasických knihoven

Obecný úvod

Identifikátory

Metadata

Interoperabilita

## Výhody digitálních knihoven:

- vzdálený a nepřetržitý přístup k informacím
- digitální data zabírají minimum prostoru
- jednodušší vyhledávání informací, fulltextové hledání
- současný přístup několika uživatelů k jednomu dokumentu
- menší pravděpodobnost zničení, poškození či ztráty dokumentů
- možnost okamžité aktualizace



# Příklady projektů digitálních knihoven

Obecný úvod

Identifikátory

Metadata

Interoperabilita





Obecný úvod

Identifikátory

Metadata

Interoperabilita

## Kahn-Wilenského architektura

- digitální objekt
  - obsah
  - metadata
  
- repozitáře
  - jednoznačné globální jméno
  - umožňuje řízený přístup k objektům
  - 2 typy metadatových záznamů
    - záznam vlastností objektu
    - transakční záznam
  - RAP (Repository Access Protocol)
  
- handle-system



Obecný úvod

**Identifikátory**

Metadata

Interoperabilita

- 1 Obecný úvod
- 2 Identifikátory**
- 3 Metadata
- 4 Interoperabilita



Obecný úvod

Identifikátory

Metadata

Interoperabilita

- koncept URN
  - globální rozsah
  - perzistence
  - škálovatelnost
  - legacy support
  - rozšiřitelnost
- klasické bibliografické identifikátory
  - ISBN
  - ISSN
  - ISADN, ISTC, ISAN, ISWC, SICI, BICI, ...
- PURL
- handle-system
- DOI



## ISBN (International Standard Book Number)

- přidělují nakladatelé
- složený identifikátor
- ISBN-10
  - ISBN 80-204-0105-9
    - identifikace země
    - identifikace nakladatele
    - konkrétní vydání
    - kontrolní číslice
  
- ISBN-13
  - + tříznakový prefix (978, 979)
  - kontrolní mechanismus
    - váhy střídavě 1, 3
    - modulo 10



## ISBN (International Standard Book Number)

- přidělují nakladatelé
- složený identifikátor
- ISBN-10
  - ISBN 80-204-0105-9
    - identifikace země
    - identifikace nakladatele
    - konkrétní vydání
    - kontrolní číslice
    - $10 \cdot 8 + 8 \cdot 2 + 6 \cdot 4 + 4 \cdot 1 + 2 \cdot 5 + 1 \cdot 9 = 143 = 13 \cdot 11$
- ISBN-13
  - + tříznakový prefix (978, 979)
  - kontrolní mechanismus
    - váhy střídavě 1, 3
    - modulo 10



Obecný úvod

**Identifikátory**

Metadata

Interoperabilita

## ISSN (International Standard Serial Number)

- přiděluje mezinárodní centrum pro ISSN
- jednoduchý identifikátor
- 1214-4029
- poslední číslice kontrolní





Obecný úvod

Identifikátory

Metadata

Interoperabilita

## ISSN (International Standard Serial Number)

- přiděluje mezinárodní centrum pro ISSN
- jednoduchý identifikátor
- 1214-4029
- poslední číslice kontrolní
- $8*1 + 7*2 + 6*1 + 5*4 + 4*4 + 3*0 + 2*2 + 1*9 = 77 = 7*11$



Obecný úvod

Identifikátory

Metadata

Interoperabilita

## URL poskytující nepřímou adresaci

- <http://purl.oclc.org/catalog/item1>
- na odkazované adrese je uloženo skutečné URL
- přesměrování pomocí http redirect
- libovolný počet PURL serverů



Obecný úvod

Identifikátory

Metadata

Interoperabilita

- kompatibilní s konceptem URN
- hdl:cnri.dlib/magazine
  - pojmenovávající autorita
  - jedinečný řetězec znaků v rámci autority
- dvojúrovňová architektura
  - globální registr
  - libovolný počet lokálních serverů
- distribuovaný systém s decentralizovanou administrací
- nutnost instalovat plugin do prohlížeče, jinak neúplná činnost



- směrovací mechanismus na základě DOI
- <http://dx.doi.org/10.1000/182>
  - prefix: 10. následovaný číslem registrující organizace
  - identifikátor digitálního objektu - jedinečný v rámci autority
- silně centralizovaný
  - povinná registrace u jediné registrační autority
- není bezplatný



Obecný úvod

Identifikátory

**Metadata**

Interoperabilita

- 1 Obecný úvod
- 2 Identifikátory
- 3 Metadata**
- 4 Interoperabilita



Obecný úvod

Identifikátory

**Metadata**

Interoperabilita

Metadata členíme do tří kategorií:

- metadata popisná
- metadata strukturální
- metadata administrativní

V praxi nemusí být toto rozdělení striktně dodržováno a jednotlivé typy se často překrývají.



## MARC (Machine Readable Cataloging)

Obecný úvod

Identifikátory

Metadata

Interoperabilita

```
01042nam a2200325 a 4500
001 cpk20031253496
003 CZ PrNK
005 20080118124007.0
007 tu
008 030610s2003 xr a e 001 0 cze
015 $a cnb001253496
020 $a 80-7302-050-5 (váz.)
035 $a (OCoLC)56868300
040 $a ABA001 $b cze
072 7 $a 004.4/.6 $x Programové vybavení. Programové prostředky $2 Konspekt $9 23
080 $a 004.451.9Linux $2 MRF
080 $a (035) $2 MRF
100 1 $a Brandejs, Michal $7 jx20070925011 $4 aut
245 10 $a Linux : $b praktický průvodce / $c Michal Brandejs
250 $a 2. vyd., (V Konvoji 1.)
260 $a Brno : $b Konvoj, $c 2003
300 $a 304 s. : $b il. ; $c 24 cm
504 $a Obsahuje bibliografii a rejstřík
650 07 $a Linux $2 czenas
650 07 $a operační systémy $2 czenas
650 09 $a Linux (operating system) $2 eczenas
655 7 $a příručky $2 czenas
655 9 $a handbooks, manuals, etc. $2 eczenas
901 $b 9878073020505 $f 2. vyd., 1. vyd. v nakl. Konvoj $o 20040506
998 $a 001253496
```



Rozmezí	Význam	Rozmezí	Význam
000-008	kontrolní pole	4XX	údaje o edici
01X-04X	kódované údaje	5XX	poznámky
05X-08X	klasifikace, signatury	6XX	věcný popis
1XX	hlavní záhlaví	70X-75X	vedlejší záhlaví
20X-24X	názvové údaje	76X-78X	vazebná pole
250-270	nakladatelské údaje	80X-830	vedl. záhlaví pro edice
3XX	fyzický popis	841-88X	holdings, lokace

Table: Struktura pole dat formátu MARC 21





- nekvalifikovaný Dublin Core
  - 15 základních metadatových prvků
    - obsah (název, předmět, popis, pokrytí, typ, zdroj, vztah)
    - intelektuální vlastnictví (tvůrce, přispěvatel, vydavatel, práva)
    - instance síťového zdroje (identifikátor, datum, jazyk, formát)
- kvalifikovaný Dublin Core
  - kvalifikátor prvku
  - kvalifikátor hodnoty
  - kvalifikátory musí splňovat princip "dumb-down"



Obecný úvod

Identifikátory

Metadata

Interoperabilita

## ■ Ukázka kvalifikovaného Dublin Core

IDENTIFIER=<http://www.ukoln.ac.uk/metadata/resources/dc/datamodel/WD-dc-rdf>: URL  
TITLE = Guidance on expressing the Dublin Core within the RDF  
TITLE = Dublin Core in RDF: Eine Anleitung  
CREATOR = Eric Miller  
CREATOR = Paul Miller  
CREATOR.Illustrator = Dan Brickley  
DESCRIPTION.Abstract = This work describes work carried out by ...  
SUBJECT.Keywords = Dublin Core; DC; Resource Description Framework; RDF; XML  
PUBLISHER = Dublin Core Metadata Initiative  
CONTRIBUTOR = Dublin Core Data Model Working Group  
DATE.Created = 1999-07-01 : ISO8601  
DATE.Revised = 1999-11-10 : ISO8601  
LANGUAGE = en : RFC1766  
TYPE = Working Draft  
FORMAT.Medium = text/html : IMT  
MYELEMENT.Checksum = 123456 : XYZ



## MODS (Metadata Object Description Schema)

- navrženo odborníky z Kongresové knihovny (Library of Congress) a Marc Standard Office
- založen na jazyce XML (titleInfo, name, typeOfResource, genre, originInfo, language, physicalDescription, abstract, tableOfContents, targetAudience, note, subject, classification, relatedItem, identifier, location, accessCondition, part, extension, recordInfo)
- obsahuje 20 základních prvků
  - dělí se na podprvky
  - mohou obsahovat atributy



# Metadata Kongresové knihovny

Obecný úvod

Identifikátory

Metadata

Interoperabilita

```
- <mods version="3.0" xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
- <titleInfo>
  <title>Sound and fury :</title>
  <subTitle>the making of the punditocracy </subTitle>
</titleInfo>
- <name type="personal">
  <namePart>Alterman, Eric</namePart>
- <role>
  <roleTerm type="text">creator</roleTerm>
</role>
</name>
<typeOfResource>text</typeOfResource>
<genre authority="marcgt">bibliography</genre>
- <originInfo>
- <place>
  <placeTerm authority="marccountry" type="code">nyu</placeTerm>
</place>
- <place>
  <placeTerm type="text">Ithaca, N.Y</placeTerm>
</place>
<publisher>Cornell University Press</publisher>
<dateIssued>c1999</dateIssued>
<dateIssued encoding="marc">1999</dateIssued>
<issuance>monographic</issuance>
</originInfo>
- <language>
  <languageTerm authority="iso639-2b" type="code">eng</languageTerm>
</language>
```



## ■ HTML

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.Title" content="Access Rights in Enterprise Full-text Search" />
<meta name="DC.Creator" content="KASPRZAK, Jan" />
<meta name="DC.Subject" content="Document-level security" />
<meta name="DC.Subject" content="Full-text search" />
<meta name="DC.Publisher" content="INSTICC" />
<meta name="DC.Date" content="2010" />
<meta name="DC.Identifier" content="https://is.muni.cz/publication/890027" />
<meta name="DC.Identifier" content="978-989-8425-04-1" />
```

## ■ RDF

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcq="http://purl.org/dc/documents/rec/dcmes-qualifiers-20000711.htm">
  <rdf:Description rdf:about="Access Rights in Enterprise Full-text Search">
    <dc>Title>Access Rights in Enterprise Full-text Search</dc>Title>
    <dc:Creator>KASPRZAK, Jan</dc:Creator>
    <dc:Subject>Document-level security</dc:Subject>
    <dc:Subject>Full-text search</dc:Subject>
    <dc:Publisher>INSTICC</dc:Publisher>
    <dc>Date>2010</dc>Date>
    <dc>Type>Text</dc>Type>
    <dc:Identifier>https://is.muni.cz/publication/890027</dc:Identifier>
    <dc:Identifier>978-989-8425-04-1</dc:Identifier>
  </rdf:Description>
</rdf:RDF>
```



Obecný úvod

Identifikátory

Metadata

**Interoperabilita**

- 1 Obecný úvod
- 2 Identifikátory
- 3 Metadata
- 4 Interoperabilita**



Obecný úvod

Identifikátory

Metadata

Interoperabilita

## Interoperabilita

*Schopnost vzájemně si rozumět, spolupracovat a dosáhnout součinnosti mezi nezávislými, technicky různorodými systémy.*



# Úroveň interoperability?

Obecný úvod

Identifikátory

Metadata

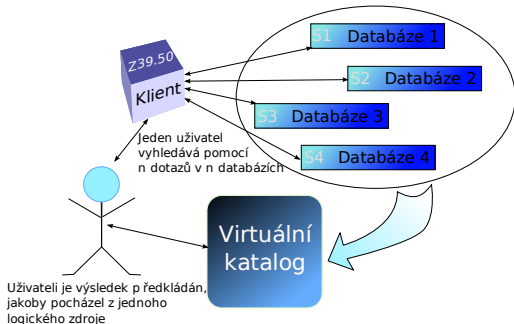
Interoperabilita

- **federace** (striktní použití standardů)
- **sklizení metadat** (digitální knihovny nabízejí základní metadata prostřednictvím jednoduchého protokolu)
- **shromažďování dat** (digitální knihovny nekooperují; informace se musí hledat explicitně pomocí služeb)
- **vyhledávací middleware** (zdroje vybavené metadaty jsou volně zapojovány do kooperace)





- architektura klient/server
- stavový protokol
- komunikaci inicializuje klient, provede sekvenci interakcí a spojení uzavře





# Open Archives Initiative (OAI)

Obecný úvod

Identifikátory

Metadata

Interoperabilita

## OAI-PMH (OAI Protocol for Metadata Harvesting)

- protokol pro hromadné **sklizení** metadat
- založen na základních otevřených standardech a protokolech (HTML, XML, HTTP)
- 6 příkazů kódovaných do URL
  - Identify
  - ListSets
  - ListIdentifiers
  - ListMetadataFormats
  - GetRecord
  - ListRecords
- příklad dotazu:  
`http://memory.loc.gov/cgi-bin/oai2_0?verb=ListIdentifiers&set=gmd&metadataPrefix=oai_dc`



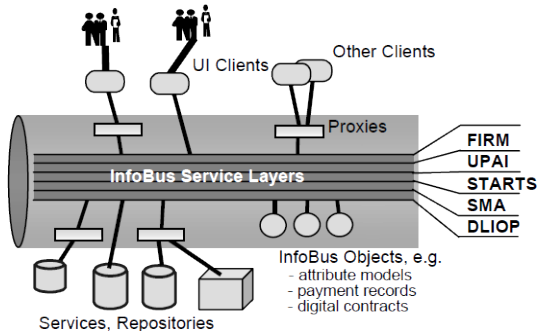
Obecný úvod

Identifikátory

Metadata

Interoperabilita

- wrapper pro každý systém



- wrappery transformují zprávy z/do interního rozhraní (protokol DLIOP)



## Open URL

- standard pro kódování metadat o zdroji do jeho URL (ANSI norma Z39.88)
- `http://resolver.example.edu/cgi?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:book&rft.isbn=0836218310&rft.btitle=The+Far+Side+Gallery+3`

## SFX

- aplikační rámec nad OpenURL



Obecný úvod

Identifikátory

Metadata

**Interoperabilita**

Děkuji za pozornost.