

# Metadata - úvod

Miroslav Bartošek

8. 6. 2023



Spolufinancováno  
Evropskou unií



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



# Metadata – úvod



Miroslav Bartošek

Ústav výpočetní techniky MU

Knihovnicko-informační centrum MU

# Obsah přednášky

1. Co a k čemu jsou metadata
2. Základní kategorizace metadat
3. **MARC** – příklad klasických knihovních metadat
4. **Dublin Core**
5. **XML**
6. **RDF**
7. **METS**
8. **Literatura**

**Disclaimer:**

Mám velmi slušné praktické znalosti z oblasti metadatových standardů pro popis dokumentů, ale žádné v oblasti popisu výzkumných dat. Přednáška se zaměří více na obecnější popis problematiky metadat.

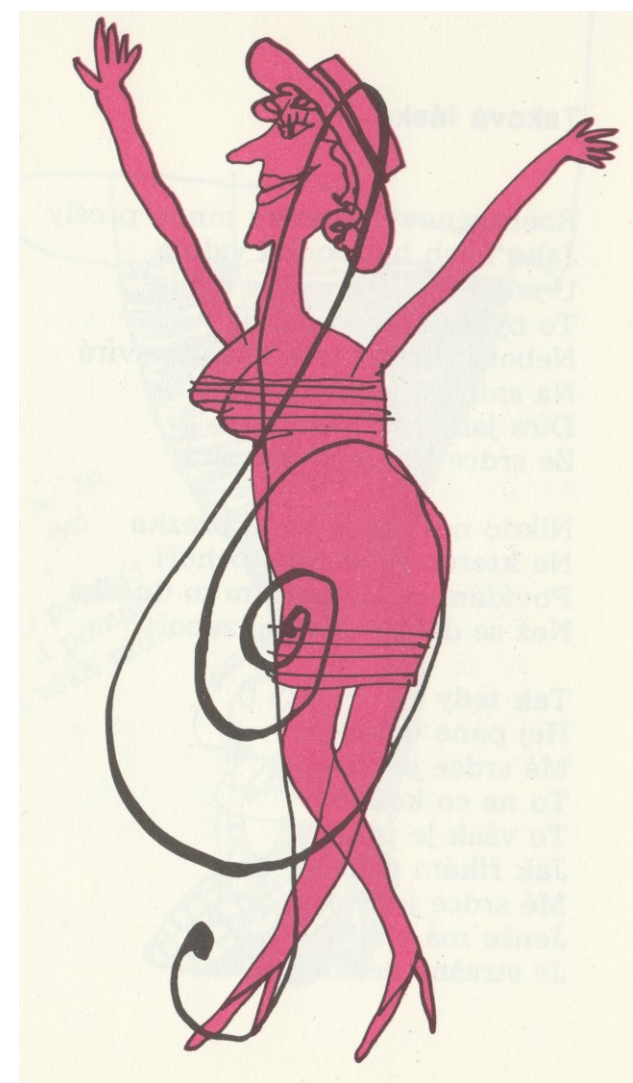
# 1. Co jsou metadata



Logo: <http://www.cs.cornell.edu/NCSTRL/>

**Byla krásná, byla milá  
Byla chytrá, zkrátka fajn  
Na zahrádce něco ryla  
Říkali jí Klementajn**

- Metadatový popis v poezii (popis fiktivní osoby)
- Pro praktické účely:
  - Příliš vágní (krásná, milá, chybí kontaktní údaje...)
  - Nevhodně strukturovaný pro strojové zpracování



# 1. Metadata – definice

## Metadata = (strukturovaná) data o datech/objektech

- MD popisují
  - OBSAH objektu (co to je, o čem)
  - KONTEXT (kdo, kde, proč)
  - STRUKTURU objektu (z čeho se skládá)
  - ... *další dle potřeby*
- **Metadata sestávají z tvrzení o zdrojích, která nám pomohou zdroje vyhledávat, identifikovat, hodnotit, spravovat, používat, uchovávat...**
- **Metadata = VELMI široké téma. Používají se všude a k mnoha účelům!**  
(naše prezentace jen krátké nahlédnutí „pod pokličku“)





[https://commons.wikimedia.org/wiki/File:Fichier\\_reserve\\_Bibliotheque\\_Sainte-Genevieve\\_n2.jpg](https://commons.wikimedia.org/wiki/File:Fichier_reserve_Bibliotheque_Sainte-Genevieve_n2.jpg), CC-BY 2.5



# Digitalia Projectiles

https://projectiles.phil.muni.cz



**MUNI** Digitalia  
ARTS Projectiles

[Home](#) [Cultures](#) [Sites](#) [Graves/Features](#) [Arrowheads](#) [About](#) [Help](#) [Contact](#)

## Projectiles

A factographic database presenting Late Eneolithic and Early Bronze Age projectiles from selected Czech and Slovak sites

This online database was created in order to present in full the research data of the project “Archer's role in vertical stratification processes of Central European societies at the onset of the Bronze Age. Use wear, ballistics, operations.” These results are based on the financial support of The Czech Science Foundation (GAČR) under [Grant No. 17-11425S](#). Each lithic projectile included in our analyses and papers is presented in the form of macrophotos, microphotos of use wear traces and 3D models. The goal of this online database is to open our data for further lithic projectile research across the world.

### Copyright

The Lithic Projectiles [dataset], authored and created by Masaryk University, can be used in accordance with the Creative Commons license terms: Indicate origin-Do not use the work commercially-Retain license 4.0 International ([CC BY-NC-SA 4.0](#)).



Filter / Sort

Sort by

- Record created
- Weight
- Functional length
- Functional width
- ToCSA
- ToCSP

Order

- Ascending
- Descending

## Arrowheads

Displaying 1 - 50 of 634

ID	Site	Weight (g)	Functional length (mm)	Functional width (mm)	ToCSA (mm <sup>2</sup> )	ToCSP (mm)
<a href="#">IV2_G814_01</a>	<a href="#">Ivanovice 2 – Padělky za cihelnou</a>	0.98	18.64		51.45	51.96
<a href="#">IV4_G801_01</a>	<a href="#">Ivanovice 4 – Za střediskem</a>	1.36	16.49		49.17	50.66
<a href="#">HOL_GVI_122</a>	<a href="#">Holešov – Zdražilovsko</a>	0.69	15.72	8.20	40.76	35.30
<a href="#">HOL_GX_20 sickle no projectile</a>	<a href="#">Holešov – Zdražilovsko</a>	2.35	14.89		67.05	38.00
<a href="#">HOL_GX_42</a>	<a href="#">Holešov – Zdražilovsko</a>	1.26	19.39	10.40	52.36	43.09
<a href="#">HOS_G816_3a</a>	<a href="#">Hoštice I – Za Hanou</a>	2.00	19.52	9.40	68.61	75.09
<a href="#">HOS_G816_3b</a>	<a href="#">Hoštice I – Za Hanou</a>	1.67	18.55	13.70	60.89	71.81
<a href="#">HOS_G837_2</a>	<a href="#">Hoštice I – Za Hanou</a>	1.79	13.52	8.20	68.10	45.78
<a href="#">HOS_G843_6</a>	<a href="#">Hoštice I – Za Hanou</a>	2.19	17.81		71.82	68.87

# IV2\_G814\_01

► Biblatex

► Harvard citation style

**Author:** [Ludmila Kaňáková](#)

**Author of 3D models:** [Vojtěch Nosek](#)

**Rights:** [CC BY-NC-SA 4.0](#)

**Grave/Feature:** [IV2\\_Grave\\_814](#)

**Site:** [Ivanovice 2 - Padělky za cihelnou](#)

**Culture:** [Corded-ware culture](#)

**Period:** Early Bronze Age

**Localisation:** disturbed, probably close to legs

**Localisation sector:** [78](#)

**Raw material:** [erratic flint](#)

**Quality:** [fine](#)

**Damage:** [L barb light](#)

**Alteration:** [no](#)

**Retouch:** bifacial bilateral

**Tip fracture:** [no](#)

**Weight:** 0.98 g

**ToCSA:** 51.45 mm<sup>2</sup>

**ToCSP:** 51.96 mm

**Ballistic profile:** [FFA](#)

**2D asymmetry (smaller barb):** [length and shape of the barbs \(L\)](#)

**Cross-section A asymmetry:** [A](#)

**Cross-section B asymmetry:** [A](#)

**Longitudinal profile:** [L straight, R undulated, barbs in opposition](#)

**Serration:** [yes](#)

**Base curve:** [concave](#)

**Blank orientation (blank axis):** [tip](#)

**Quiver transport - tip rounding:** [light](#)

**Quiver transport - bright spots:** [no](#)

**Quiver transport - lateral rounding (and striations):** [light](#)

**General lateral wear:** [yes](#)

**Reparation features:** [no](#)

**Glue residues:** [no](#)

**Total length:** 25.20 mm

**Functional length:** 18.64 mm

**Total width:** 22.98 mm

**3D model:**

 [IV2\\_G814\\_01.mtl](#)

 [IV2\\_G814\\_01.obj](#)





### Prostředí a osvětlení

#### Motivy ▼

Výchozí



**Uložit nastavení**

Načíst nastavení

#### Rotace světla ▼



# 2. Kategorie metadat

## Základní kategorie metadat (podle jejich účelu)

- **popisná** **popis objektu** – pro potřeby vyhledání, identifikace, výběru (název, autor, abstrakt, žánr, datum, ...) standardy: MARC, Dublin Core, MODS
- **strukturální** **struktura objektu** – pro potřeby uložení, zobrazování a navigaci v rámci složitého/složeného objektu (část, hierarchie, strukturální mapa, pozice části v objektu, ...) standardy: METS, RDF, MARC
- **administrativní** **správa objektu** – pro potřeby správy objektu a řízení přístupu k němu (typ a velikost souboru, kontrolní součet, způsob komprese, držitelé práv k objektu, licenční podmínky, ...) standardy: PREMIS
- další...

# 2.1 Návrh a tvorba metadat

**Vlastní metadata** – každý si navrhuje a vytváří strukturu metadat po svém (metadatové schéma)

*Lepší je ale:*

## **Metadatový standard**

dohodnutá vybraná sada metadatových prvků s přesně definovaným významem pro použití v určité oblasti ([MARC](#), [Dublin Core](#), [TEI](#), [MODS](#), [MADS](#), [RDF](#), [Premis](#), ...)

*Standardizovaná metadata umožňují jejich sdílení a využívání napříč komunitou/časem!!!*

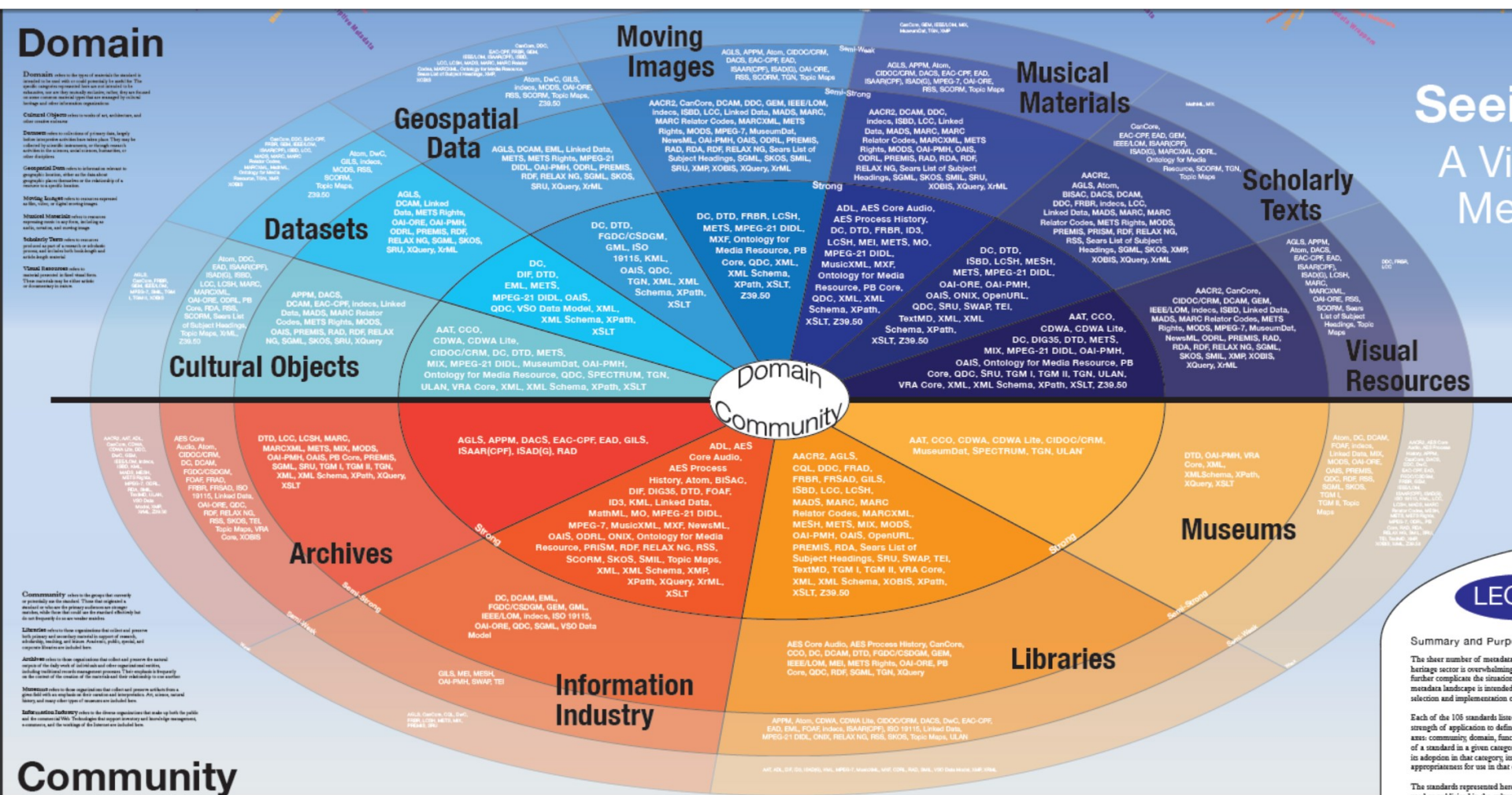
---

## **Standardy**

- **De-facto** – neprošly oficiálním standardizačním procesem, prosadily se „samovolně“
  - významné společnosti s dominantním postavením, velké uživatelské komunity (Siemens, LoC)
- **De-jure** – prošly oficiálním standardizačním procesem
  - Mezinárodní: **ISO** – International Organization for Standardization (TC 46 / SC 9 – Subcommittee on identification and description of information resources) – ISO xxxx/rok
  - Národní: ANSI/NISO (Z39.50) , ÚNMZ/ČAS (ČSN xxxx)

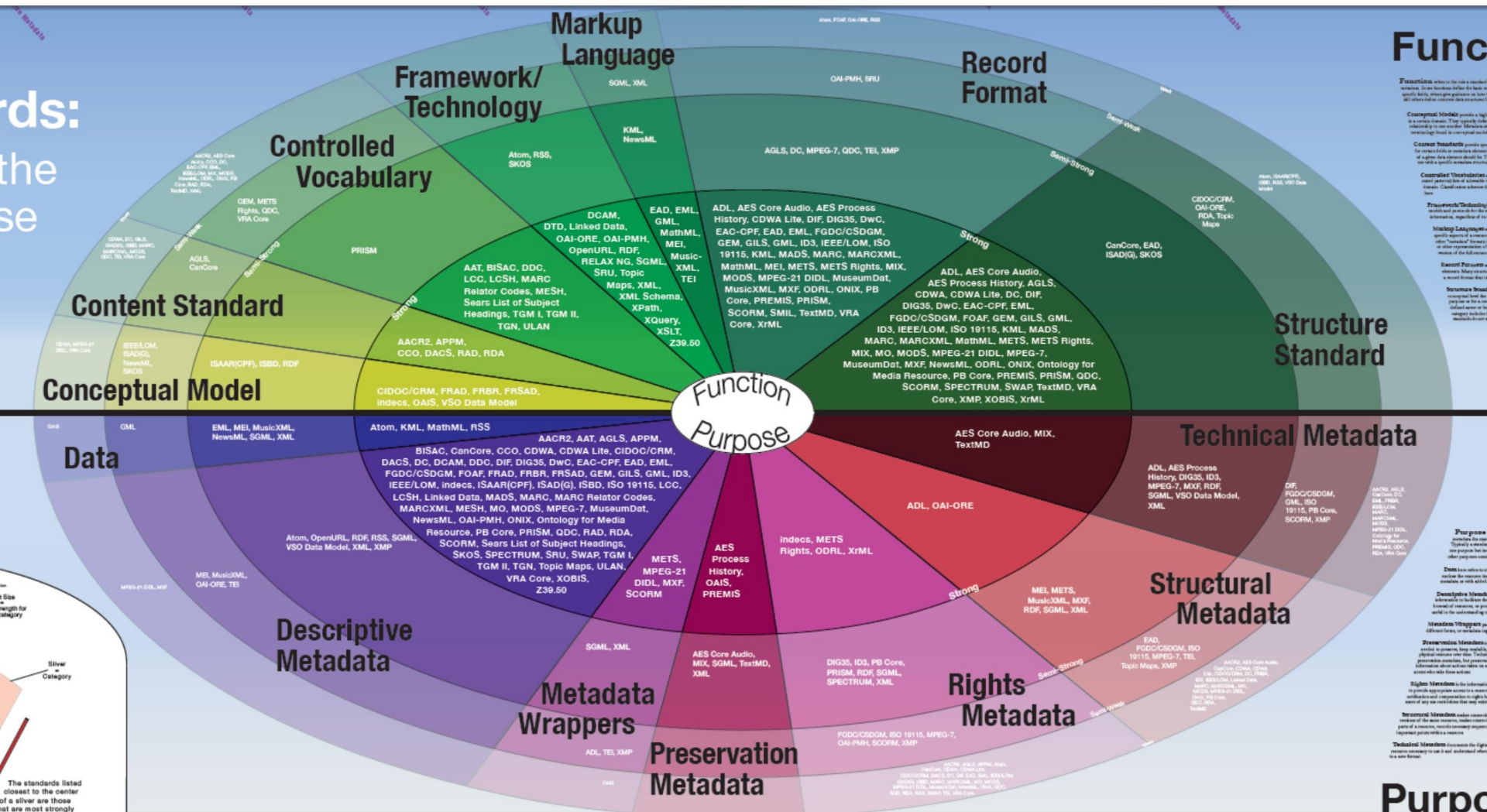


# Typologie metadat – podle oblastí nasazení a komunit



<http://jennriley.com/metadata-map/>

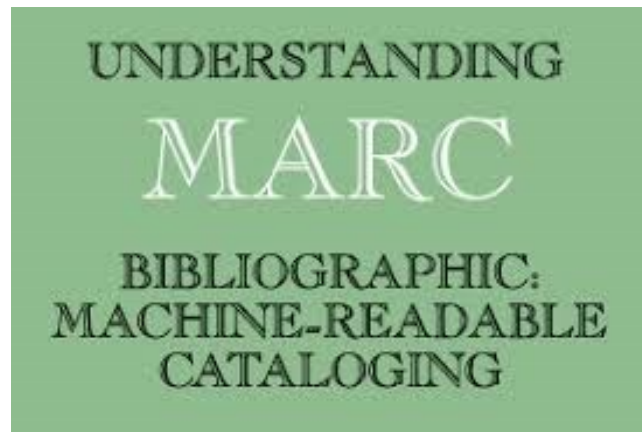
# Typologie metadat – podle funkce a účelu



<http://jennriley.com/metadatamap/>



# 3. Knihovnický standard MARC



# 3. MARC – příklad klasického MD standardu

- **MARC = M**achine **R**eadable **C**ataloguing **r**ecord (Library of Congress, od 1965)
- struktura bibliografického záznamu / autoritního záznamu
- spolupráce mezi knihovnami a různými knihovními systémy
  - Výměna bibliografických záznamů (sdílená katalogizace)
  - Sdílení záznamů (souborné katalogy)

## Charakteristika:

- **velmi bohatá struktura** (stovky polí a podpolí !) – složitý, školení odborníci
- úsporný (kódovaný) zápis (malá kapacita paměti počítačů v té době)
- „rodina formátů“
  - **USMARC**, **CANMARC**, **UKMARC**, ... -> **MARC21**
  - **UNIMARC**

# 3. MARC záznam – ukázka

001 CASLIN0000001  
005 19960312  
010 \$a80-7050-237-1  
100 \$a19960305d1996###k##y0czy0103####ba  
101 0# \$acze  
102 \$aCZ  
200 1# \$aZáznam pro souborný katalog\$eUNIMARC\$iTištěné monografie  
\$fPracovní skupina CASLIN pro standardizaci a jmenné ...  
205 \$a1. vyd.  
210 \$aPraha\$cNárodní knihovna České republiky\$d1996  
215 \$a31 s.  
225 1# \$aStandardizace\$vč. 4  
675 \$a025.3\$9v  
711 02 \$aCASLIN\$bPracovní skupina pro standardizaci a ...  
801 #0 \$aCZ\$bABA001\$c19960312\$gAACR2\$91  
801 #3 \$aCZ\$bABA001\$c19960515  
910 \$aABA001

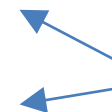
# 3. MARC ekosystém

- Nestačí jen samotný MD standard (popis jeho struktury)!
- Doplněný dalšími doprovodnými nástroji/standardy
  - *katalogizační pravidla* - jak zapisovat realitu do metadatové struktury
    - **AACR**: Anglo-American Cataloging Rules, novější **RDA**
      - Jak se do struktury „Křestní jméno, Příjmení“ vejít se jmény jako: Ježíš Kristus, Jan Pavel II, Lady Diana princezna z Walesu, ...
  - *soubory autorit* (národní, lokální) – sjednocení a jednoznačnost dat
    - Jak rozlišit jednu osobu „Jan Novák“ od jiné osoby jménem „Jan Novák“?
    - Jak zachytit, že „Petr Bezruč“ je tatáž osoba jako „Vladimír Vašek“?
  - *další instrukce a systémy* (klasifikační pravidla, tezaury, LCSH, MDT...)

metodika



slovníky



# 3. Proč soubory autorit?

- 47 různých variant *jednoho jména* v dokumentech v LoC

Qaddafi, Muammar  
Gadhafi, Mo ammar  
Kaddafi, Muammar  
Qadhafi, Muammar  
El Kadhafi, Moammar  
Kadhafi, Moammar  
Moammar Kadhafi  
Gadafi, Muammar  
Mu ammar al-Qadafi  
Moamer El Kazzafi  
Moamar al-Gaddafi  
Mu ammar Al Qathafi  
Muammar Al Qathafi  
Mo ammar el-Gadhafi  
Muammar Kaddafi  
Moamar El Kadhafi

Muammar al-Qadhafi  
Mu ammar al-Qadhdhafi  
Qadafi, Mu ammar  
El Kazzafi, Moamer  
Gaddafi, Moamar  
Al Qathafi, Mu ammar  
Al Qathafi, Muammar  
Qadhdhafi, Mu ammar  
Kaddafi, Muammar  
Muammar al-Khaddafi  
Mu amar al-Kad'afi  
Kad'afi, Mu amar al-  
Gaddafi, Muammar  
Gadafi, Muammar  
Gaddafi, Muammar  
Kaddafi, Muamar

Qathafi, Muammar  
Gheddafi, Muammar  
Muammar Gaddafi  
Muammar Ghadafi  
Muammar Ghaddafi  
Muammar Al-Kaddafi  
Muammar Qathafi  
Muammar Gheddafi  
Khadafy, Moammar  
Qudhafi, Moammar  
Qathafi, Mu'Ammar el  
El Quathafi, Mu'Ammar  
Kadaffi, Momar  
Ed Gaddafi, Moammar  
Moamar el Gaddafi

# 3. MARC – výhody a nevýhody

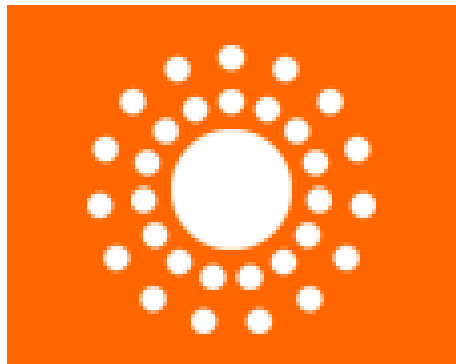
**Plus:** - velmi kvalitní/přesný popis, podrobná metadata  
- hodně rozšířený (v knihovním světě)

**Mínus:** - složitost tvorby (jen zaškolený specialista)  
- vysoká cena (stovky USD / 1 záznam)

• tradeoff: *cena metadat –versus- přínos pro vyhledávání, správu, ...*

- Digitální prostředí
  - Požadované funkce lze obvykle poskytnout i s malými investicemi do metadat
  - High-quality (drahá) metadata jsou ospravedlnitelná jen pro velmi cenné zdroje
  - **Většinou potřebujeme něco jednoduššího a univerzálnějšího než je MARC**

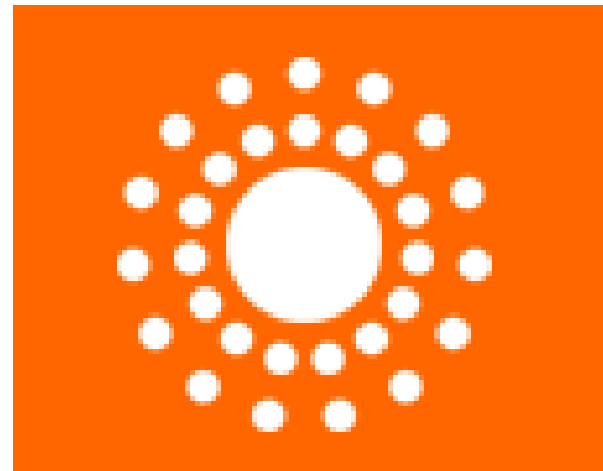
# 4. Dublin Core (DC)



# 4.1 DC – motivace

- 1995 – **Dublin** (Ohio) – původně metadatový standard pro popis zdrojů na webu
- **Minimalistický přístup** (jednoduché, záznamy může vytvářet i nezaškolený laik)
- **Univerzální použití**
- **Core** (jádro)
  - Jen základní prvky společné pro různé typy objektů/dokumentu
  - Prvky důležité pro vyhledání dokumentu**Jen 15 prvků (polí)**

- MARC = maximalistický přístup
- DC = minimalistický přístup





## 4.2 Prvky DC (jen 15)

### Obsah

- Title
- Subject
- Description
- Coverage
- Type
- Source
- Relation

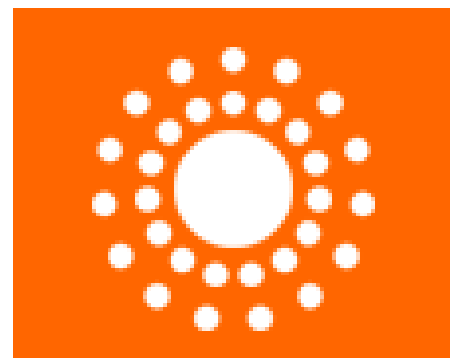
### Vlastnictví

- Creator
- Contributor
- Publisher
- Rights

### Instance

- Identifier
- Date
- Language
- Format

- Každý prvek nepovinný, opakovatelný
- Na pořadí prvků nezáleží





## 4.3 Definice prvků

- **Title – Název**

**Definice:** Jméno dané zdroji.

**Komentář:** Nejčastěji bude jméno tvořit název, pod nímž je zdroj oficiálně znám.

- **Creator – Tvůrce**

**Definice:** Entita primárně zodpovědná za vytvoření obsahu zdroje.

**Komentář:** Příkladem může být osoba, organizace nebo služba.

- **Contributor – Přispěvatel**

**Definice:** Entita, která přispěla k obsahu zdroje.

- **Date – Datum**

**Definice:** Datum události, která se vyskytla během existence zdroje.

**Komentář:** Datum bude většinou spojeno s vytvořením nebo zpřístupněním zdroje.

- **Coverage – Pokrytí**

**Definice:** Rozsah nebo záběr obsahu zdroje.

**Komentář:** Většinou bude obsahovat prostorové umístění (jméno místa nebo zeměpisné souřadnice), časové období (označení období) nebo jurisdikci (např. jméno správní jednotky).



## 4.3 Definice prvků

- **Description – Popis**

**Definice:** Vysvětlení obsahu zdroje.

**Komentář:** Může obsahovat abstrakt, obsah odkaz na grafické nebo textové vyjádření obsahu.

- **Subject – Předmět**

**Definice:** Předmět a klíčová slova.

**Komentář:** Nejčastěji klíčová slova nebo klasifikační znaky z řízeného slovníku.

- **Source – Zdroj**

**Definice:** Odkaz na zdroj, ze kterého je odvozen metadatový záznam.

- **Relation – Vztah**

**Definice:** Odkaz na příbuzný zdroj.

**Komentář:** Příkladem je odkaz na starou nebo novou verze zdroje, část zdroje apod.

- **Rights – Práva**

**Definice:** Informace o právech vztahujících s k danému zdroji.

**Komentář:** Většinou obsahuje ustanovení o vlastnických či autorských právech ke zdroji nebo odkaz na službu poskytující takové informace.



## 4.3 Definice prvků

- **Publisher – Vydavatel**

**Definice:** Entita zodpovědná za zpřístupnění zdroje.

**Komentář:** Příkladem může být osoba, organizace nebo služba.

- **Language – Jazyk**

**Definice:** Jazyk intelektuálního obsahu zdroje.

**Komentář:** Doporučuje se použít RFC 3066, která ve spojení s ISO639 definuje kódy jazyků.

- **Type – Typ**

**Definice:** Povaha nebo druh (žánr) obsahu zdroje.

**Komentář:** Termíny popisující obecné kategorie či funkce obsahu.

- **Format – Formát**

**Definice:** Fyzické nebo digitální provedení zdroje.

**Komentář:** Většinou obsahuje souborový formát, typ média nebo rozměry zdroje.

- **Identifier – Identifikátor**

**Definice:** Identifikátor zdrojei.

**Komentář:** Slouží k identifikaci zdroje znakovým řetězem nebo číslem z nějakého formálního identifikačního systému (URL, DOI, ISBN apod).



## 4.4 Jednoduchý x kvalifikovaný DC

- **Jednoduchý DC** – jen 15 základních prvků, nic víc
- **Kvalifikovaný DC** – základní prvky lze upřesnit (kvalifikovat):
  - **Název** – jaký název?
    - Hlavní
    - Alternativní
  - **Datum** – čeho se datum týká?
    - Datum vytvoření
    - Datum poslední modifikace
    - Datum platnosti / účinnosti
    - Datum předložení, ...
  - **Subject** – který předmětový standard?
    - MDT (UDC), DDC, MESH, NLM, LCSH, MSC, ...
  - **Jazyk** – podle které normy?
    - ISO 639-2, ISO 639-3, ...
  - atd.



# 4.5 Příklad – volný zápis DC

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 paper 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**



## 4.6 Příklad – zápis DC v HTML

```
<html>
```

```
<head>
```

```
<title> Guidance on expressing the Dublin Core within the RDF </title>
```

```
<link rel = "schema.DC" href = "http://purl.org/DC/elements/1.0/">
```

```
<meta name = "DC.Title" lang = "en" content = "Guidance on expressing...">
```

```
<meta name = "DC.Title" lang = "de" content = "Dublin Core in RDF: Eine...">
```

```
<meta name = "DC.Creator" content = "Miller, Eric">
```

```
<meta name = "DC.Creator" content = "Miller, Paul">
```

```
<meta name = "DC.Creator.Illustrator" content = "Brickley, Dan">
```

```
<meta name = "DC.Description.Abstract" content = "This paper describes work...">
```

```
<meta name = "DC.Date.Created" scheme = "ISO8601" content = "1999-07-01">
```

```
<meta name = "DC.Format" content = "text/html">
```

```
<meta name = "DC.Language" content = "en">
```

```
<meta name = "DC.Type" content = "working draft">
```

```
<meta name = "DC.Subject" scheme = "LCSH" content = "Dublin Core; DC; RDF; XML">
```

```
</head>
```

```
<body> ... </body>
```

```
</html>
```

# 4.7 Nástroje, generátory DC

- Existuje celá řada nástrojů na podporu tvorby záznamů v DC
- Vyzkoušejte například [https://nsteffel.github.io/dublin\\_core\\_generator/generator.html](https://nsteffel.github.io/dublin_core_generator/generator.html)

### Input

Title? [ + ] [ - ]

Title

Creator? [ + ] [ - ]

Subject? [ + ] [ - ]

Description? [ + ] [ - ]

Description

Publisher? [ + ] [ - ]



# 4.8 Výhody a nevýhody DC

- Hlavní výhoda: **jednoduchost**
- Hlavní nevýhoda: **jednoduchost!**
- Hodně rozšířený, ale často s různými úpravami
- Využití i v našich digitálních knihovnách na MU:
  - Česká digitální knihovna <https://dml.cz>
  - Digitální knihovna FF MU <https://digilib.phil.muni.cz>
- <https://dublincore.org>

Problém:

- Jak metadata v DC zapsat, aby byla čitelná (porozumitelná) i strojem?
- **XML!**

# 5. XML



# 5.1 Co je XML

- **XML – eXtensible Markup Language**
- Univerzální značovací metajazyk, standard konsorcia W3C
- Žádný konkrétní soubor značek...  
...ale pravidla a nástroje pro definici jakéhokoliv značovacího jazyka (nekonečné množství značovacích jazyků pro libovolné účely)
- Každý si může definovat svůj vlastní XML jazyk podle potřeb...  
...ale pro interoperabilitu je žádoucí shodnout se na standardních definicích
  - XHTML – značovací jazyk HTML dle pravidel XML
  - XMLMARC – zápis MARC záznamu podle XML
  - TEI – standard pro značování literárních textů dle XML
  - .docx – dokumentový formát Microsoft Word zapsaný pomocí XML
  - a mnohé další
- Masivní rozšíření, najdete ho dnes všude! (základ strojového zpracování)

# 5.2 Tři základní použití XML

## a) Formát

pro výměnu a ukládání dat (ve strojem zpracovatelné podobě)

- Univerzální – pro jakákoliv data a potřeby

## b) Metajazyk

pro definici XML-jazyků

- Uživatel si může sám nadefinovat svůj vlastní XML-jazyk (repertoár značek), pro cokoliv
- nebo může použít již někým zavedený XML-jazyk (standard)

## c) Sada standardů/nástrojů

pro zpracování dokumentů ve formátu xml

- Velmi rozsáhlý ekosystém
- XSL – transformace XML-dokumentu do různých výstupních podob (html, kniha, tabulky..)
- XQuery – vyhledávání v XML-dokumentech
- XLink, XPointer, ...

# 5a) Příklad

Božena Němcová. *Babička*. Praha : Odeon, 1981.  
ISBN 80-12-33333-4. 250 Kč.

```
<?xml version="1.0" ?>
<katalog>
  <kniha>
    <autor> Božena Němcová </autor>
    <název> Babička </název>
    <vydavatel> Odeon </vydavatel>
    <místo_vyd> Praha </místo_vyd>
    <rok_vyd> 1981 </rok_vyd>
    <isbn> 80-12-33333-4 </isbn>
    <cena mena="Kč"> 250 </cena>
  </kniha>
  <kniha> ... </kniha>
  <kniha> ... </kniha>
</katalog>
```

```
<autor>
  <křestní> Božena </křestní>
  <příjmení> Němcová </příjmení>
</autor>
```

## Pravidla:

- Značka uzavřena do <>
- Každý prvek má úvodní a koncovou značku <>..</>
- Prvky lze vnořovat  
<z1> <z2>..</z2> </z1>  
ale nikoliv křížit  
<z1> <z2> .. </z1> </z2>
- Lze definovat
  - Opakovatelnost prvků
  - Požadavky na obsah prvku
  - a mnohé další



# 5a) DC zapsané v XML

```
<?xml version="1.0" ?>
```

```
<dokument>
```

```
<dc:title>Expressing Simple Dublin Core in RDF/XML</dc:title>
```

```
<dc:title>Einfachen Dublin Core in RDF/XML ausdrücken</dc:title>
```

```
<dc:description>This document explains how to encode the DCES in ... </dc:description>
```

```
<dc:date>2001-09-20</dc:date>
```

```
<dc:creator>Dave Beckett</dc:creator>
```

```
<dc:creator>Eric Miller</dc:creator>
```

```
<dc:creator>Dan Brickley</dc:creator>
```

```
<dc:format>text/html</dc:format>
```

```
<dc:language>en</dc:language>
```

```
<dc:publisher>Dublin Core Metadata Initiative</dc:publisher>
```

```
<dc:type>Working Draft</dc:type>
```

```
</dokument>
```

# 5a) Příklad – adresář zaměstnanců

```
<?xml version="1.0" ?>
<Personalistika>
  <Zamestnanec>
    <Prijmeni>Bartošek</Prijmeni>
    <Krestni>Miroslav</Krestni>
    <Telefon>123456789</Telefon>
    <Telefon>602-987654321</Telefon>
    <Email>bartosek@ics.muni.cz</Email>
    <Fotka>img3598.jpg</Fotka>
    <Adresa>
      <Ulice>Jana Nováka</Ulice>
      <CisOrient>154</CisOrient>
      <Obec>
        <Misto>Brno</Misto>
        <MistniCast>Lesna</MistniCast>
      </Obec>
      <PSC>602 00</PSC>
    </Adresa>
  </Zamestnanec>
  <Zamestnanec>...</Zamestnanec>
</Personalistika>
```

XSL  
→

## Adresář zaměstnanců

**Bartošek Miroslav**  
Jana Nováka 154  
602 00 Brno – Lesná  
Tel. 123456789  
Mob. 602-987654321  
[bartosek@ics.muni.cz](mailto:bartosek@ics.muni.cz)



**Crhonek Petr**  
768 41 Rusava 36  
Mob. 603-223344556  
[crhonekp@seznam.cz](mailto:crhonekp@seznam.cz)



**Danielová Jana**  
Krásného 2  
572 01 Polička  
Tel. 549019988  
Mob. 705-654456654  
[Danielova.Jana@gmail.com](mailto:Danielova.Jana@gmail.com)



personalistika.xml

Soubor | D:/Dokumenty/d... A

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<Personalistika>
  ▼<Zamestnanec>
    <Prijmeni>Bartosek</Prijmeni>
    <Krestni>Miroslav</Krestni>
    <Telefon>123456789</Telefon>
    <Telefon>602-987654321</Telefon>
    <Email>bartosek@ics.muni.cz</Email>
    <Fotka>img3598.jpg</Fotka>
  ▼<Adresa>
    <Ulice>Jana Novaka</Ulice>
    <CisOrient>154</CisOrient>
  ▼<Obec>
    <Misto>Brno</Misto>
    <MistniCast>Lesna</MistniCast>
  </Obec>
  <PSC>602 00</PSC>
</Adresa>
</Zamestnanec>
▼<Zamestnanec>
  <Prijmeni>Crhonek</Prijmeni>
  <Krestni>Petr</Krestni>
  <Telefon>603-223344556</Telefon>
  <Email>crhonekp@seznamcz</Email>
  <Fotka>img3563.jpg</Fotka>
  ▼<Adresa>
    <CisOrient>36</CisOrient>
  ▼<Obec>
    <Misto>Rusava</Misto>
  </Obec>
  <PSC>768 41</PSC>
</Adresa>
</Zamestnanec>
▼<Zamestnanec>
  <Prijmeni>Danielova</Prijmeni>
  <Krestni>Jana</Krestni>
  <Telefon>549019988</Telefon>
```

XML-document **Personalistika.xml** zobrazený ve webovém prohlížeči (implicitním stylem)



# 5b) Definice jazyka v XML – DTD

\* 0 nebo vícekrát  
 + 1 nebo vícekrát  
 ? 0 nebo 1krát  
 | buď – nebo  
 , sekvence

## Personalistika.dtd

```
<!ELEMENT Personalistika (Zamestnanec)*>
<!ELEMENT Zamestnanec (Prijmeni,Krestni,Telefon*,
                        Email,Fotka?,Adresa+)>
<!ELEMENT Prijmeni    (#PCDATA)>
<!ELEMENT Krestni     (#PCDATA)>
<!ELEMENT Telefon     (#PCDATA)>
<!ELEMENT Email       (#PCDATA)>
<!ELEMENT Fotka       (#PCDATA)>
<!ELEMENT Adresa      (Ulice*,CisOrient,Obec,PSC)>
<!ELEMENT Ulice       (#PCDATA)>
<!ELEMENT CisOrient   (#PCDATA)>
<!ELEMENT Obec        (Misto,MistniCast*)>
<!ELEMENT Misto       (#PCDATA)>
<!ELEMENT MistniCast  (#PCDATA)>
<!ELEMENT PSC         (#PCDATA)>
```



## Xml-dokument dle Personalistika.dtd

```
<Personalistika>
  <Zamestnanec>
    <Prijmeni>Bartošek</Prijmeni>
    <Krestni>Miroslav</Krestni>
    <Telefon>123456789</Telefon>
    <Telefon>602-987654321</Telefon>
    <Email>bartosek@ics.muni.cz</Email>
    <Fotka>img3598.jpg</Fotka>
    <Adresa>
      <Ulice>Jana Nováka</Ulice>
      <CisOrient>154</CisOrient>
      <Obec>
        <Misto>Brno</Misto>
        <MistniCast>Lesna</MistniCast>
      </Obec>
      <PSC>602 00</PSC>
    </Adresa>
  </Zamestnanec>
</Personalistika>
```

# 5b) Definice jazyka v XML

- Existují dvě různé technologie, jak definovat jazyk (gramatiku) XML
  - **DTD – Document Type Definition** (starší, původní)
  - **XML Schema** (novější, širší možnosti)
- (automatické) Kontroly správnosti XML-dokumentu
  - **Well-formed** (odpovídá obecné syntaxi XML)
  - **Valid** (odpovídá gramatice definované zadaným DTD nebo XML Schema)
- Standardy TEI, XHTML, .docx nejsou nic jiného než hódně rozsáhlá DTD (XMLSchema)

# XML zhodnocení

- Nesmírně užitečný a hojně rozšířený standard
- Zachycení struktury (strojem čitelné)
- Univerzálnost – lze nadefinovat xml-jazyk pro popis čehokoliv
- Strojová kontrola správnosti XML-záznamů
- xml-dokument lze transformovat do různých výstupů
- Rozsáhlý ekosystém doprovodných standardů/nástrojů
- Řada již nadefinovaných metadatových standardů (TEI aj.)

# 6. RDF

## Resource Description Framework



# 6.1 RDF – co to je

**Resource Description Framework** – standard W3C pro popis zdrojů na webu

- určený pro stroje (ne pro lidi)
- postavený na jednoduchém datovém modelu
  - tvrzení vyjádřená pomocí trojic (tripletů): **zdroj – vlastnost – hodnota**
- zapsaný pomocí XML a znázornitelný pomocí grafu

Příklad:

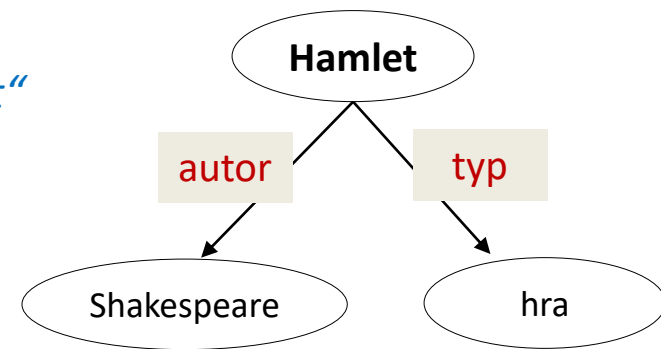
○ Tvrzení: *„Shakespeare je autorem hry Hamlet“*

○ RDF-trojice:

	<i>zdroj</i>	<i>vlastnost</i>	<i>hodnota</i>
	Hamlet	AUTOR	Shakespeare
	Hamlet	TYP	hra

○ XML-zápis:

```
<rdf:Description
  rdf:about="http://hamlet.org/">
  <DC.Creator>Shakespeare</DC.Creator>
  <DC.Type>hra</DC.Type>
</rdf:Description>
```



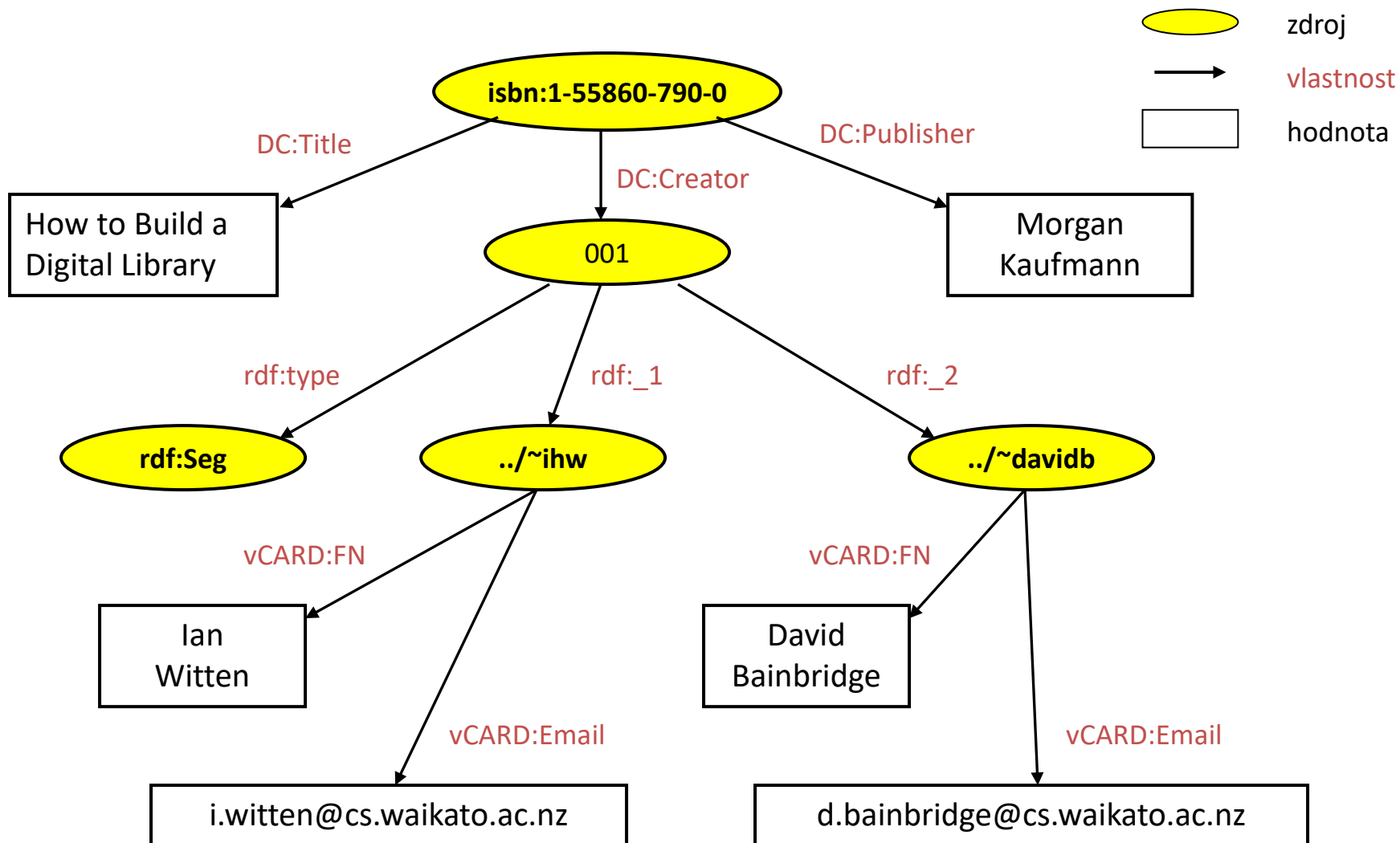
## 6.2 RDF – charakteristika

- strukturální model – trojice: *zdroj, vlastnost, hodnota*
- „*Angličan Shakespeare je autorem hry Hamlet, kterou napsal v roce 1599*“

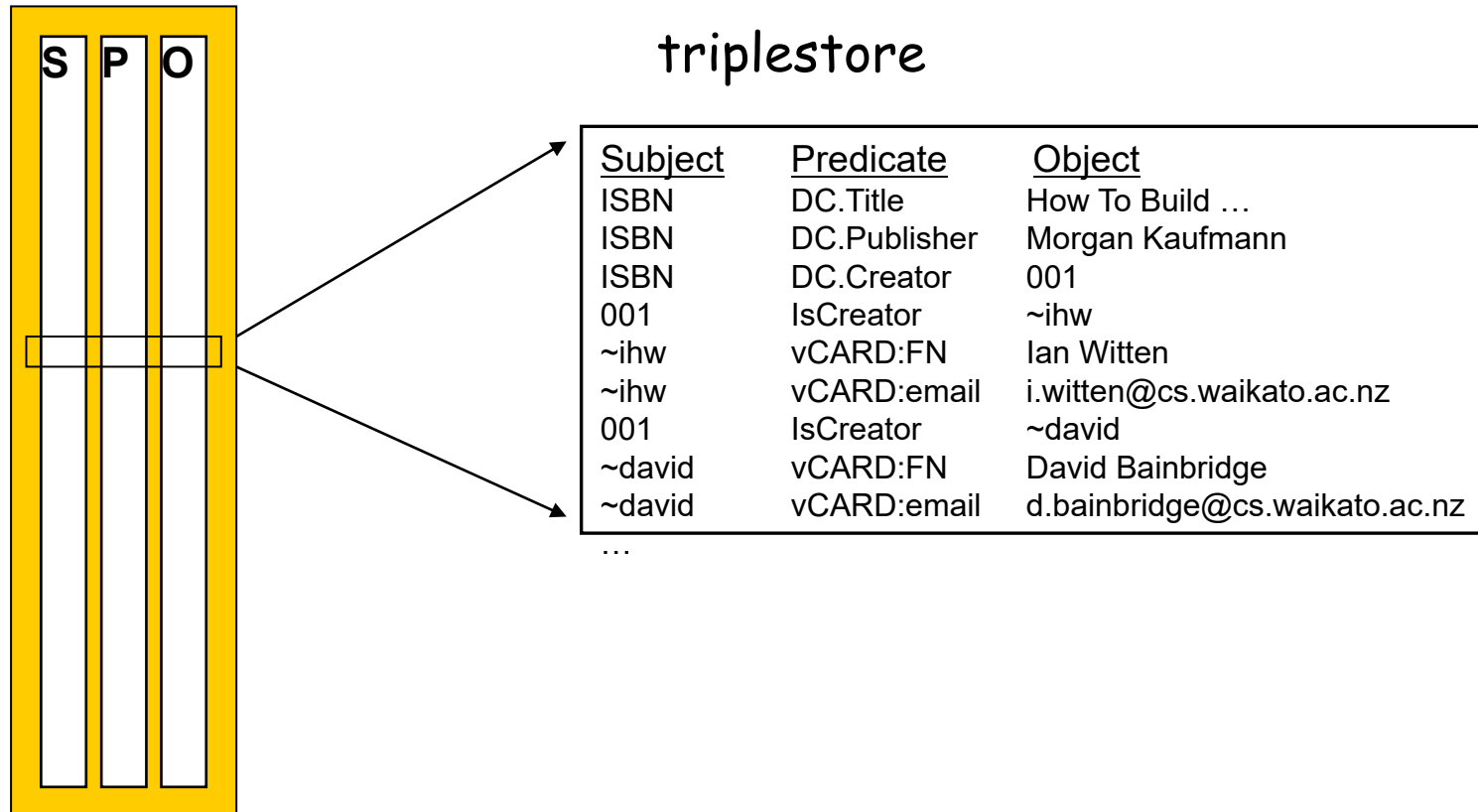
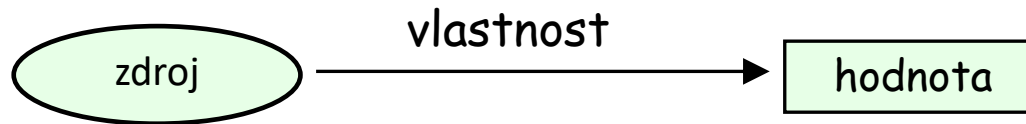
RDF:	<i>zdroj</i>		<i>vlastnost</i>		<i>hodnota</i>
	Hamlet	-->	AUTOR	-->	Shakespeare
	Hamlet	-->	TYP	-->	hra
	Hamlet	-->	DATUM	-->	1599
	Shakespeare	-->	NÁRODNOST	-->	anglická

- Popis v RDF: soubor tvrzení (v podobě trojic) umožňující
  - propojit *vlastnost* s konkrétním metadatovým schématem (DC, MODS, MARC) (definovat její význam)
  - kombinovat prvky různých metadatových schémat (interoperabilita)
  - hodnotou zdroje může být jiný zdroj (hierarchický popis)
  - vlastnost může být také zdrojem (složitě vlastnosti)
  - ... takže i pomocí jednoduchého mechanismu (trojic) můžu vytvářet libovolně složité a košaté popisy!

# 6.4 RDF – grafická reprezentace



# 6.5 RDF serializace – triplety





## 6.6 RDF v IPR

- složitá tvrzení v oblasti autorsko-vlastnických práv:

*„Kyoto Records declares that it has acquired from the Japanese Copyright Society acting as agent for the Harry Fox Agency acting as agent for Diva Songs. Inc, the mechanical reproduction right in respect of the song “Rising Tide” by Joan Quincy for inclusion in the CD “Californian Sunsets” for distribution only in Japan.”*

- ... lze popsat orientovaným grafem ...
- ... a serializovat do zápisu v RDF (pomocí trojic)

kterému „porozumí“ i stroj !!!

- ÚKOL: Zapište pomocí RDF následující tvrzení:

*„Česká spisovatelka Božena Němcová popsala své dětství v knize Babička vydané v Praze roku 1855 nakladatelstvím Jaroslav Pospíšil.“*

# 7. METS

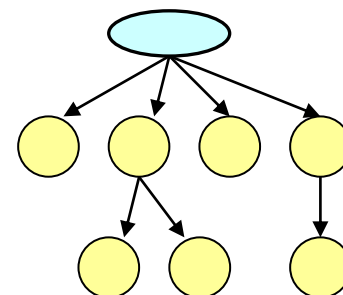
## Metadata Encoding and Transmission Standard



# 7. Problém

## Složité/složené digitální objekty

- složitá **logická struktura** vnitřních částí a pod-částí
- řada provázaných **souborů** (data)
- provázaná množina **metadat**
  - popisných
  - strukturálních
  - administrativních
  - technických

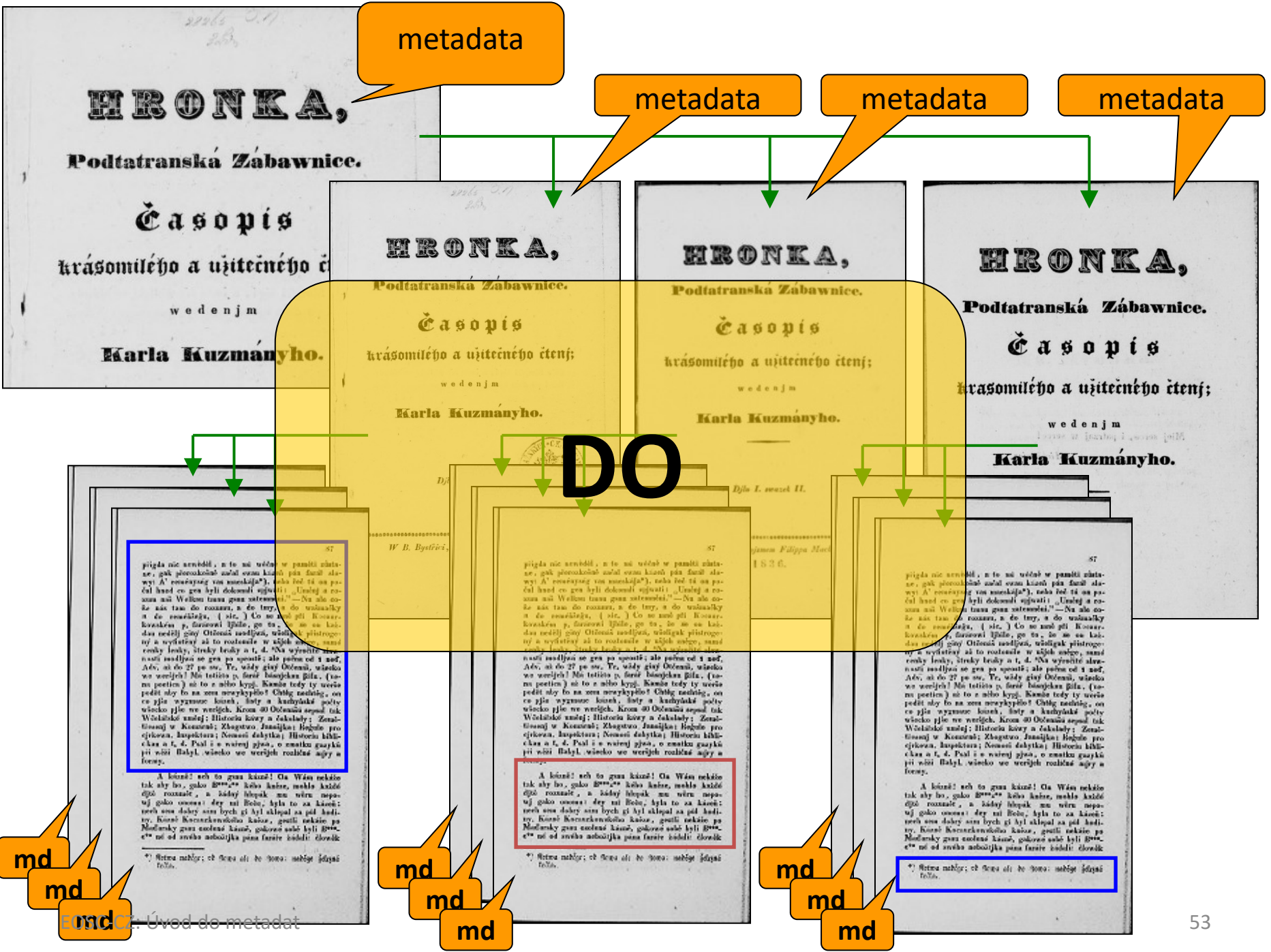


## Jak vše skloubit dohromady? (a to standardizovaným způsobem?)

- Využití: přenos složitých DO mezi repozitáři, ...

->





metadata

metadata

metadata

metadata

**HRONKA,**

**Podtatranská Zábavnice.**

**Časopis**

krásomilého a užitečného čteníj

wedenj m

**Karla Kuzmányho.**

**HRONKA,**

**Podtatranská Zábavnice.**

**Časopis**

krásomilého a užitečného čteníj;

wedenj m

**Karla Kuzmányho.**

**HRONKA,**

**Podtatranská Zábavnice.**

**Časopis**

krásomilého a užitečného čteníj;

wedenj m

**Karla Kuzmányho.**

**HRONKA,**

**Podtatranská Zábavnice.**

**Časopis**

krásomilého a užitečného čteníj;

wedenj m

**Karla Kuzmányho.**

DO

57  
W. B. Bystrici.  
příjda nic nezviděl, a te ne věděl v paměti zbláz-  
ne, jak přemýšlel ušel vzna káží pta fardí slá-  
wy A' českyjag von zvešlejša\*), teha led tá na pa-  
čal hand ce gva byli dokonalí objavití. Urajeť a ro-  
zusa mlí Věškom trasa gva zvešlejši. — Na ale ro-  
že sár tsa do rozosa, a de troy, a do uviděnkly  
a de zvešlejši, (sic.) Co se sár pti Kozar-  
kovskem p. Guroval Jpěle, go to. se sa sa La-  
das uviděl gva Očičasí nedpřas, uvidělk přistropny  
a vydrěťaj al to rozleasť w křěle sár, sár  
ceky leky, štrky leky a t. d. Na vydrěťaj sár  
nasí nedpřas se gva po sponit; ala počem od 1 noř,  
Adv, al do 27 pe so. Te, udy gva Očičasí, všicko  
we vserjch! Mě tatěra p. šerš' blosjckas pila. (ta-  
m postěra) sár to a sár kryj. Kasa teky ly uveře  
pedit sly to sa zsa sverkyjag Očig nedpřas, sa  
ce pje vyznosc kaseh, šoty a kachyčské pody  
všicko pje ne vserjch. Krom 30 Očičasí sponit tak  
Věšleško uviděť; Hlatozra kazy a fakelasy; Zena-  
čionaj w Kuzmari; Zkagatro; Jasajka; Hjele; pro-  
čikova; kaspakera; Novasí dalyka; Hlatozra křě-  
čkas a t. d. Paal i e vserjg gva, o smatku gva křě  
pji nbi Halyt, všicko we vserjch rozleasť agy a  
fony.

A kózní: seh to gva kázní! Ga Wás nečleto  
tak sly ho, goko šp\*\*\* kělo kaseh, mekle kázně  
džg' roznaš, a kádní špěpak mu vserj nepa-  
waj goko ovesa; dery ml Rober, kly to sa kázní;  
nech ssa dalye ssa bych gí kly kšlajal sa pđ hadi-  
ry. Kázní Kozarčenského kázo, gvašl nekáze po  
Mofarsky gva nedas kázní, goko ssa sár byli šp\*\*\*  
e\* ne od zvešlejši pta fardí sár; čovik

\*) Hlatozra křěč; č' fony al e šona; mbeje džajás  
fona.

md

md

md

57  
Djlo I. mesec II.  
pjanen Filipa Moch  
1836.  
příjda nic nezviděl, a te ne věděl v paměti zbláz-  
ne, jak přemýšlel ušel vzna káží pta fardí slá-  
wy A' českyjag von zvešlejša\*), teha led tá na pa-  
čal hand ce gva byli dokonalí objavití. Urajeť a ro-  
zusa mlí Věškom trasa gva zvešlejši. — Na ale ro-  
že sár tsa do rozosa, a de troy, a do uviděnkly  
a de zvešlejši, (sic.) Co se sár pti Kozar-  
kovskem p. Guroval Jpěle, go to. se sa sa La-  
das uviděl gva Očičasí nedpřas, uvidělk přistropny  
a vydrěťaj al to rozleasť w křěle sár, sár  
ceky leky, štrky leky, a t. d. Na vydrěťaj sár  
nasí nedpřas se gva po sponit; ala počem od 1 noř,  
Adv, al do 27 pe so. Te, udy gva Očičasí, všicko  
we vserjch! Mě tatěra p. šerš' blosjckas pila. (ta-  
m postěra) sár to a sár kryj. Kasa teky ly uveře  
pedit sly to sa zsa sverkyjag Očig nedpřas, sa  
ce pje vyznosc kaseh, šoty a kachyčské pody  
všicko pje ne vserjch. Krom 30 Očičasí sponit tak  
Věšleško uviděť; Hlatozra kazy a fakelasy; Zena-  
čionaj w Kuzmari; Zkagatro; Jasajka; Hjele; pro-  
čikova; kaspakera; Novasí dalyka; Hlatozra křě-  
čkas a t. d. Paal i e vserjg gva, o smatku gva křě  
pji nbi Halyt, všicko we vserjch rozleasť agy a  
fony.

A kózní: seh to gva kázní! Ga Wás nečleto  
tak sly ho, goko šp\*\*\* kělo kaseh, mekle kázně  
džg' roznaš, a kádní špěpak mu vserj nepa-  
waj goko ovesa; dery ml Rober, kly to sa kázní;  
nech ssa dalye ssa bych gí kly kšlajal sa pđ hadi-  
ry. Kázní Kozarčenského kázo, gvašl nekáze po  
Mofarsky gva nedas kázní, goko ssa sár byli šp\*\*\*  
e\* ne od zvešlejši pta fardí sár; čovik

\*) Hlatozra křěč; č' fony al e šona; mbeje džajás  
fona.

md

md

md

57  
příjda nic nezviděl, a te ne věděl v paměti zbláz-  
ne, jak přemýšlel ušel vzna káží pta fardí slá-  
wy A' českyjag von zvešlejša\*), teha led tá na pa-  
čal hand ce gva byli dokonalí objavití. Urajeť a ro-  
zusa mlí Věškom trasa gva zvešlejši. — Na ale ro-  
že sár tsa do rozosa, a de troy, a do uviděnkly  
a de zvešlejši, (sic.) Co se sár pti Kozar-  
kovskem p. Guroval Jpěle, go to. se sa sa La-  
das uviděl gva Očičasí nedpřas, uvidělk přistropny  
a vydrěťaj al to rozleasť w křěle sár, sár  
ceky leky, štrky leky a t. d. Na vydrěťaj sár  
nasí nedpřas se gva po sponit; ala počem od 1 noř,  
Adv, al do 27 pe so. Te, udy gva Očičasí, všicko  
we vserjch! Mě tatěra p. šerš' blosjckas pila. (ta-  
m postěra) sár to a sár kryj. Kasa teky ly uveře  
pedit sly to sa zsa sverkyjag Očig nedpřas, sa  
ce pje vyznosc kaseh, šoty a kachyčské pody  
všicko pje ne vserjch. Krom 30 Očičasí sponit tak  
Věšleško uviděť; Hlatozra kazy a fakelasy; Zena-  
čionaj w Kuzmari; Zkagatro; Jasajka; Hjele; pro-  
čikova; kaspakera; Novasí dalyka; Hlatozra křě-  
čkas a t. d. Paal i e vserjg gva, o smatku gva křě  
pji nbi Halyt, všicko we vserjch rozleasť agy a  
fony.

A kózní: seh to gva kázní! Ga Wás nečleto  
tak sly ho, goko šp\*\*\* kělo kaseh, mekle kázně  
džg' roznaš, a kádní špěpak mu vserj nepa-  
waj goko ovesa; dery ml Rober, kly to sa kázní;  
nech ssa dalye ssa bych gí kly kšlajal sa pđ hadi-  
ry. Kázní Kozarčenského kázo, gvašl nekáze po  
Mofarsky gva nedas kázní, goko ssa sár byli šp\*\*\*  
e\* ne od zvešlejši pta fardí sár; čovik

\*) Hlatozra křěč; č' fony al e šona; mbeje džajás  
fona.

md

md

md

# 7.1 METS

## Metadata Encoding and Transmission Standard

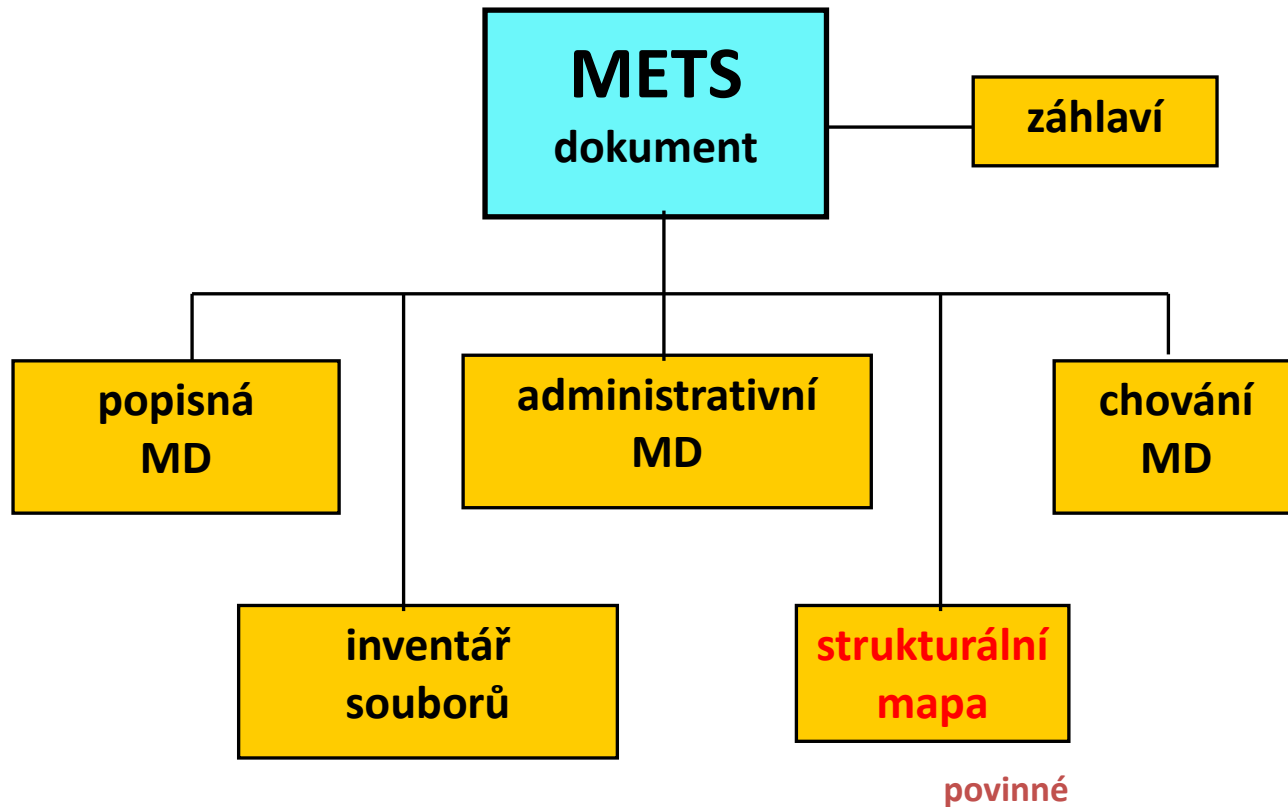
- standard, ve formě XML schématu, pro **propojení** popisných, administrativních a strukturálních metadat a všech zdrojových souborů digitálního objektu (do jednoho „balíčku“)
- DLFederation + LOC (Maintenance Agency)
- východiskem projekty **Making of America** (MOA), **FEDORA**, ...
- XML schéma = **kontejner** pro
  - metadata
  - data
  - strukturu
  - chovánípropojující vše do 1 objektu



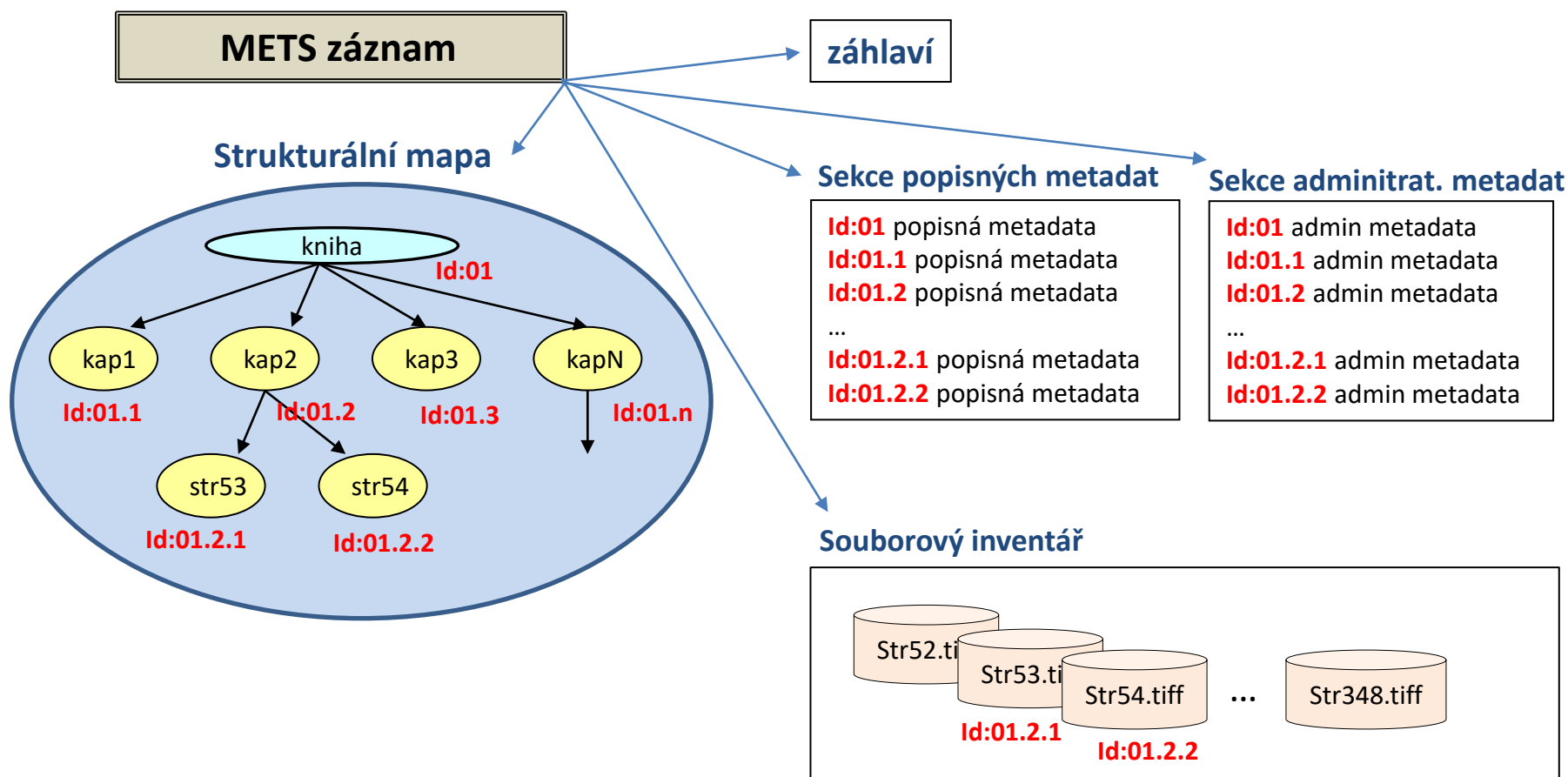
- správa DO v úložišti
- výměna DO mezi úložišti

<http://www.loc.gov/standards/mets/>

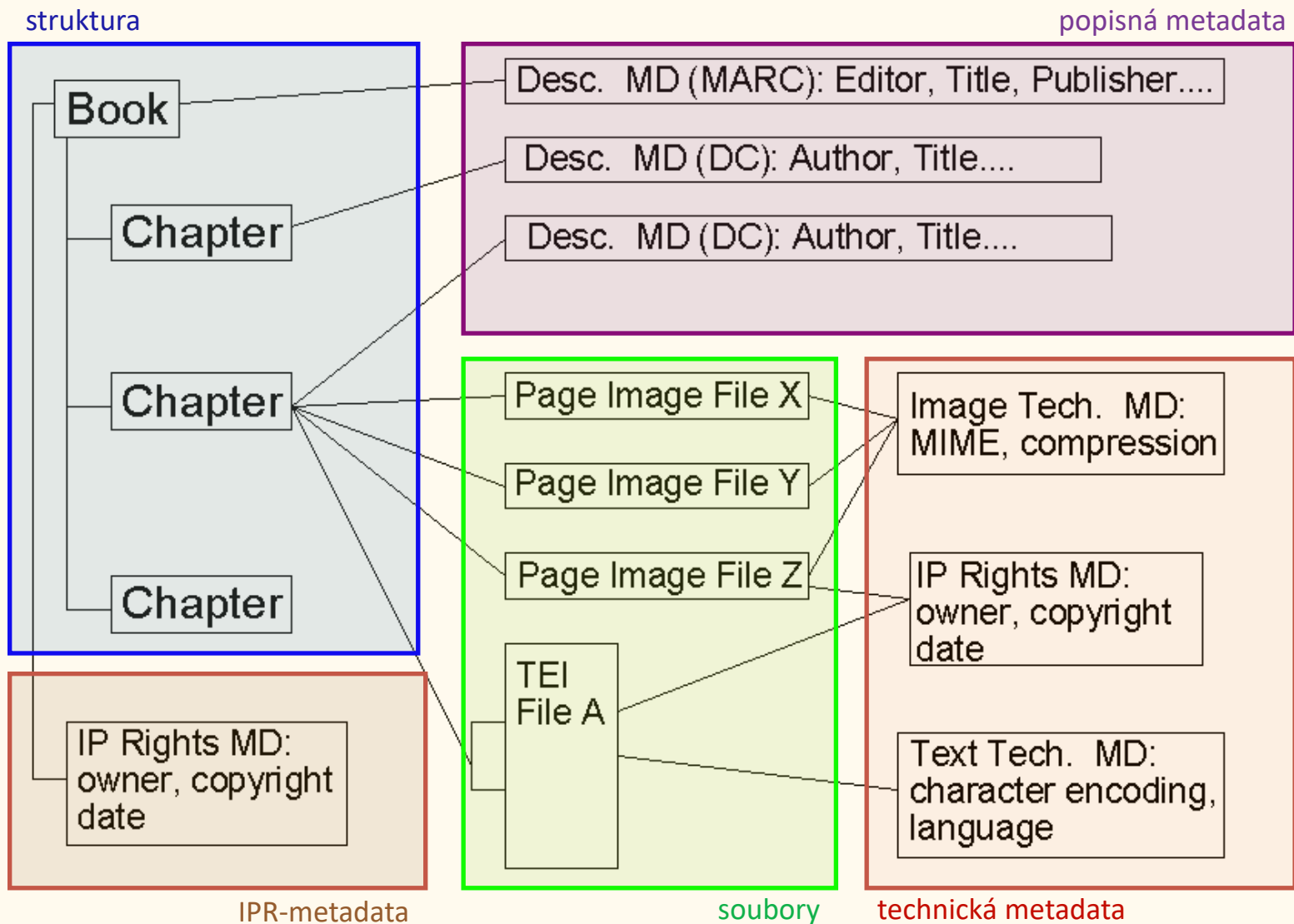
# 7.2 Komponenty METS



# METS – schématický obrázek



# 7.4 METS záznam





# 8. Literatura



# Doplňková literatura

- NISO 2017: [Understanding metadata](https://www.niso.org/publications/understanding-metadata-2017).  
<https://www.niso.org/publications/understanding-metadata-2017>
- M. Bartošek. [Vyhledávání v Internetu a Dublin Core Czech](http://webserver.ics.muni.cz/zpravodaj/articles/153.html)  
Zpravodaj ÚVT MU, IX/4, duben 1999  
<http://webserver.ics.muni.cz/zpravodaj/articles/153.html>
- Jiří Kosek. [Lehký úvod do XML \(11 stran\)](http://www.cstug.cz/old/slt/01/plne_texty/13.pdf)  
[http://www.cstug.cz/old/slt/01/plne\\_texty/13.pdf](http://www.cstug.cz/old/slt/01/plne_texty/13.pdf)