

Digitální knihovny

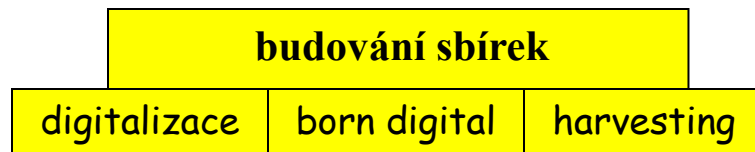
Metadata I



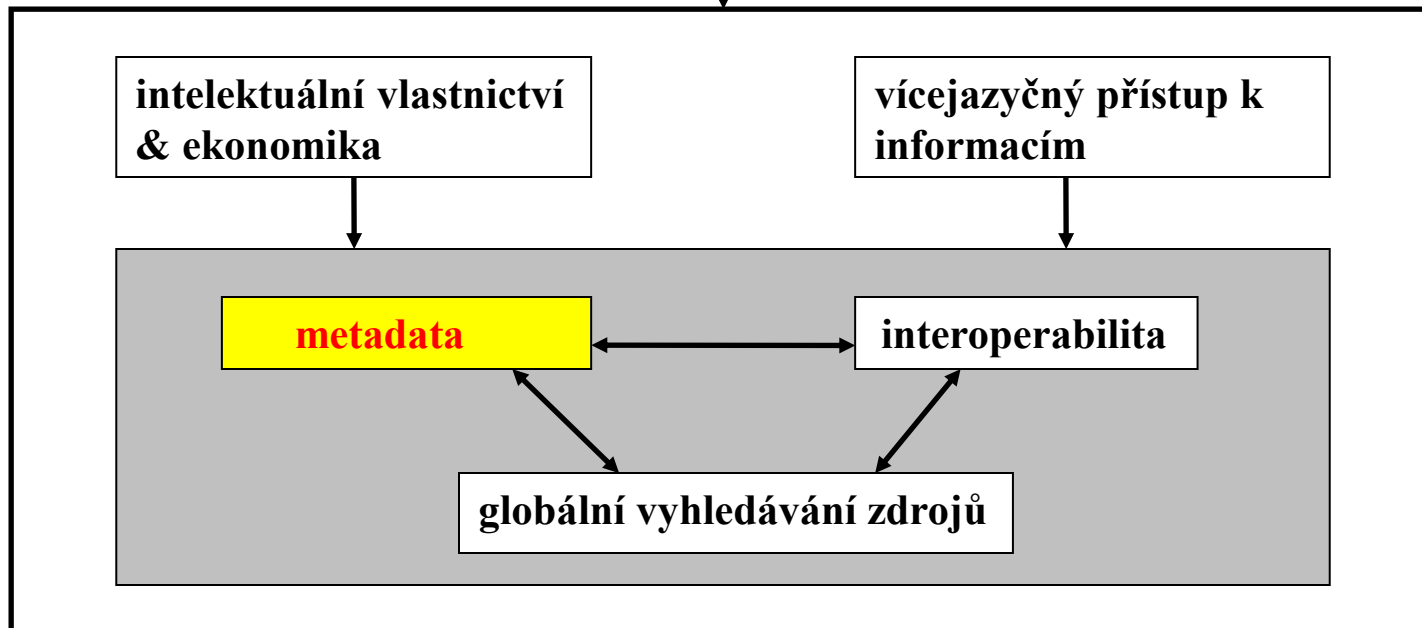
Miroslav Bartošek

Ústav výpočetní techniky MU

Knihovnicko-informační centrum MU



obecný rámec a architektura DL



zobecněný model dokumentu

dlouhodobé uchování digitální informace

Obsah přednášky

1. Úvod a přehled
 2. Knihovnická metadata – MARC
-

Síťová metadata

3. RFC 1807
4. DC – Dublin Core
5. Metadata Kongresové knihovny

1. Metadata – přehled



1. Metadata – definice

- **Metadata = (strukturovaná) data o datech**
- popisují
 - OBSAH zdroje (o čem)
 - KONTEXT (kdo, kde, proč)
 - STRUKTURU zdroje (jak)
 - ... další
- **Metadata consists of statements we make about resources to help us find, identify, use, manage, evaluate, and preserve them.**

Marty Kurth, tutorial on DC Semantics,
<http://dc2006.ucol.mx/program.htm>
- **Metadata = VELMI široké téma !**

1. Příklady

Katalogizační lístek v knihovně

- Popis knih – jmenný rejstřík, věcný rejstřík, ...

Katalog předmětů v IS MU

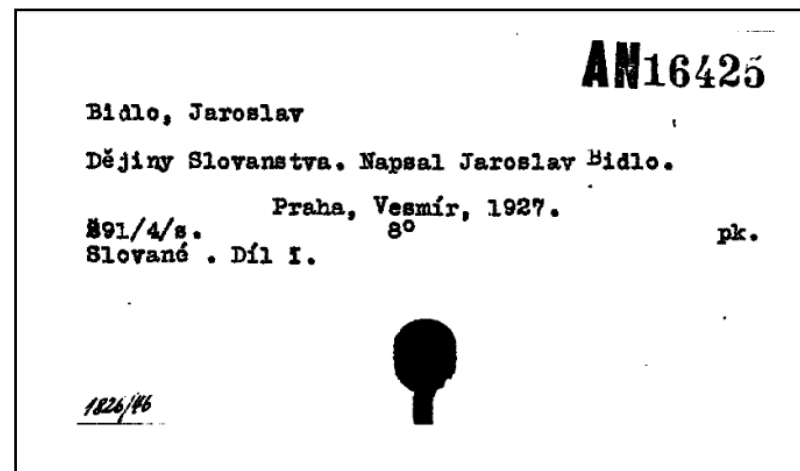
- Popis vyučovaných kurzů

Záhlaví webové stránky

- Popis webové stránky (<HEAD> ... </HEAD>)

EXIF záznam v obrázku JPEG

- Popis fotografie



ARTS020 Digitální humanitní vědy ✨

Filozofická fakulta
podzím 2020

▣ Rozsah

2/0/0. 5 kr. Ukončení: k.
Vyučováno online.

▣ Vyučující

PhDr. Michal Lorenz, Ph.D. (přednášející)

▣ Garance

PhDr. Michal Lorenz, Ph.D.
Katedra informačních studií a knihovnictví - Filozofická fakulta
Kontaktní osoba: Mgr. Alice Lukavská
Dodavatelské pracoviště: Katedra informačních studií a knihovnictví - Filozofická fakulta

▣ Rozvrh

Po 14:00–15:40 [B2.13](#)

▣ Předpoklady

! PROGRAM (B - ISK_) | (PROGRAM (B - ISK_) & (ARTS001 Úvod do kognitivních věd | | ARTS002 Přístup kultura | | ARTS004 Homér a evropská lit. | | ARTS005 Život v kyberprostoru | | ARTS006 The Fall of an | | ARTS007 Germánské literatury – nositelé Nobelovy ceny za literaturu | | ARTS008 Argumentace a logika | | ARTS009 Mongolska a stepních říší | | ARTS011 Proměny křesťanské Evropy: Víze, kritika, diskuse | | ARTS012 Ku | | ARTS013 Slované | | ARTS014 Jak číst světovou literaturu. | | ARTS015 Mezi kulturou a přírodou | | ARTS017 Život po smrti v evrop. umění | | ARTS018 | | ARTS019 Shakespeare a povaha lásky | | ARTS020 (Mám splněno? -)

▣ Omezení zápisu do předmětu

Předmět je otevřen studentům libovolného oboru.
Předmět si smí zapsat nejvýše 140 stud.
Momentální stav registrace a zápisu: zapsáno: 53/140, pouze zareg.: 1/140, pouze zareg. s předností (m

▣ Cíle předmětu

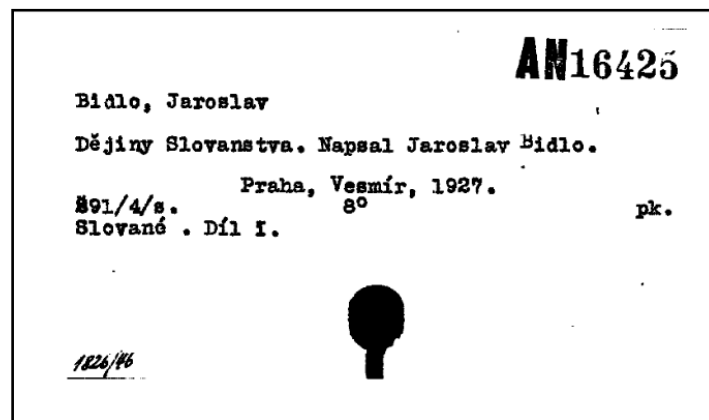


https://commons.wikimedia.org/wiki/File:Fichier_reserve_Bibliotheque_Sainte-Genevieve_n2.jpg, CC-BY 2.5

1. Knihovnická x síťová metadata

Metadata z pohledu DL

- dlouhá a „usazená“ tradice v knihovnách (katalogy, rejstříky, ...)
- bouřlivý (živelný?) vývoj v digitálním prostředí
- klasická knihovnická metadata : **kompletní popis**
- síťová metadata : **specializovaná !**
- příčiny odlišností
 - *způsob tvorby*
 - centrální x různé nezávislé komunity
 - ruční x automatizovaný
 - *způsob přístupu ke zdroji*
 - nepřímý x přímý



1.1 Typologie metadat

a) Základní typy metadat – podle jejich účelu

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

Metadatový standard (metadatové schéma)

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

1.1 Typologie metadat

b) podle oblasti nasazení

- popis/vyhledávání zdrojů (Dublin Core, MARC, MODS, RFC 1807, TEI)
- identifikace (DOI metadata, BibTeX)
- dlouhodobá archivace (PREMIS)
- IPR, e-commerce (XrML, ODRL, INDECS, ONIX, MPEG-21)
- content rating (PICS)
- interoperabilita (METS, RDF, MPEG-21 DIDL)
- formátově specializovaná (MPEG-7, TIFF, EXIF)
- oborově specializovaná (fyzika, CSDGM – GIS, medicína...)
- ...

c) podle místa uložení

- *vložená přímo do zdroje* (záhlaví HTML, JPEG-EXIF metadata, XML zdroje) – *embedded*
- *externí – asociovaná* (databáze metadat těsně svázané se zdroji)
- *externí – třetí strana* (v nezávislých DB, bez přímé kontroly zdroje, viz např. abstraktové databáze)
- *centralizovaná x distribuovaná*

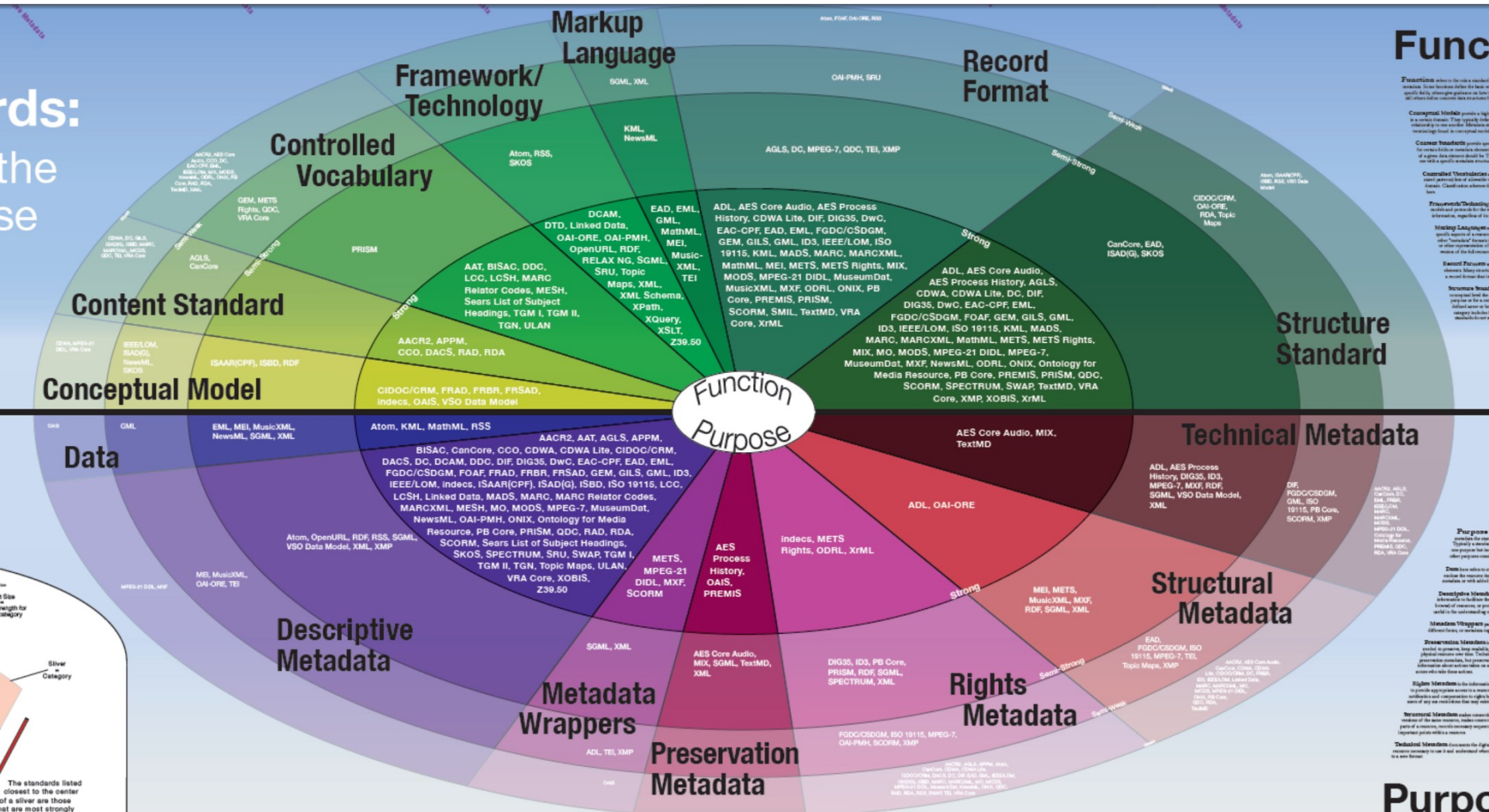
Typů a druhů metadat pro různé oblasti nasazení existuje fakt hodně...

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)



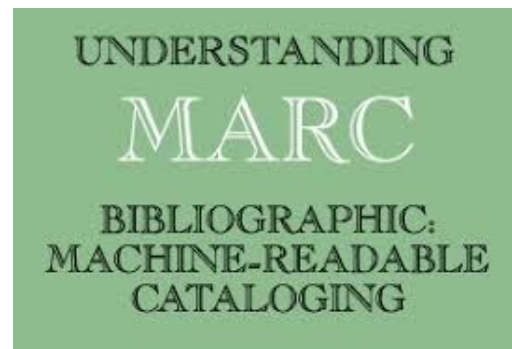
<https://xkcd.com/927/>

Typologie metadat – podle funkce a účelu (ilustrační obrázek)



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

2. Knihovnická metadata MARC

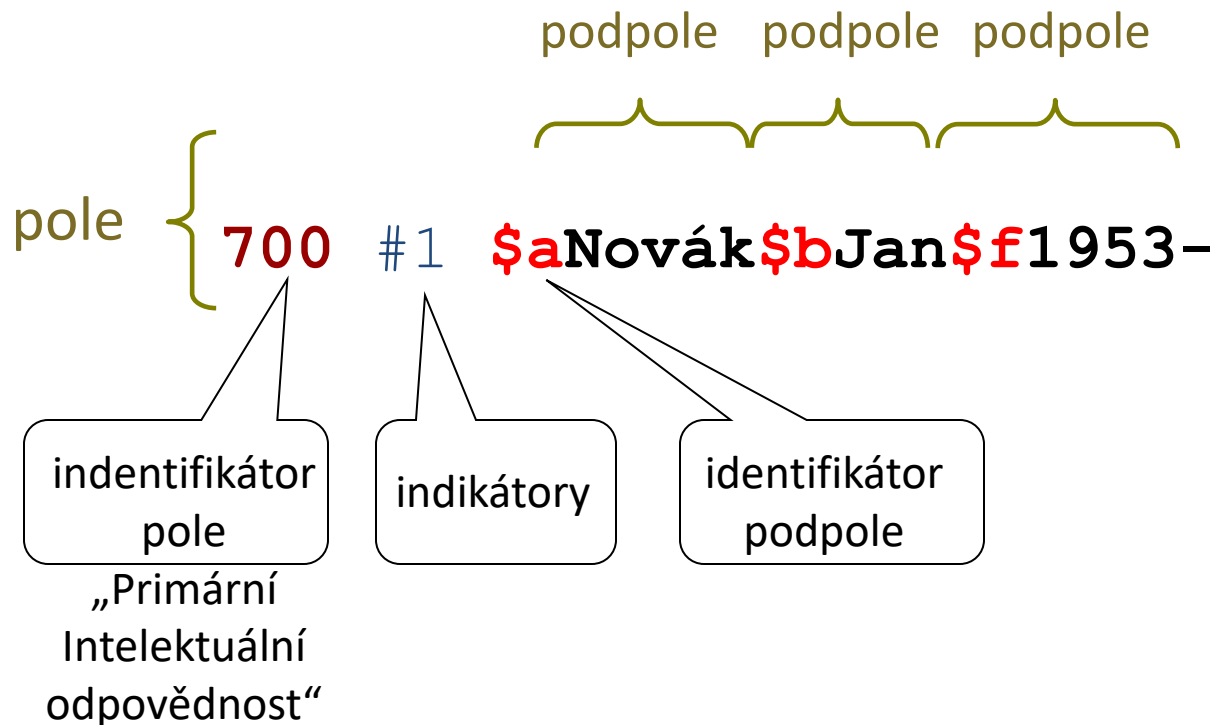


2. Rodina standardů MARC

- **MARC = M**Achine **R**eadable **C**ataloguing **r**ecord (LoC, od 1965)
- obecná *struktura* bibliografického záznamu
 - kódovací schéma pro prvky záznamu (markup)
 - výměnný formát pro přenos záznamů (transfer syntax)
- velmi rozšířený - spolupráci mezi knihovnami a různými systémy
 - souborné katalogy (union catalogues)
 - výměna (bibliografických) záznamů
- **velmi bohatá struktura** (stovky polí a podpolí !) + **úsporný zápis**
- rodina formátů typu MARC:
 - **USMARC**, **CANMARC**, **UKMARC**, ... -> **MARC21**
 - **UNIMARC** (IFLA, 1977, nejprve jako převodní můstek) → **UNIMARC-CZ (93)**

2.1 Pole podpole MARC

- MARC záznam se skládá z polí proměnné délky
- pole může být členěno do podpolí (opakování)



2.2 UNIMARC – řádkový

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

2.3 UNIMARC – přehled

- UNIMARC – **bloky polí**:
 - **0xx** blok identifikačních čísel
 - **1xx** kódovaných informací
 - **2xx** popisných informací
 - **3xx** poznámek
 - **4xx** propojovacích polí
 - **5xx** souvisejících názvů
 - **6xx** předmětové analýzy
 - **7xx** intelekt. zodpovědnosti
 - **8xx** zdrojových informací
 - **9xx** národního použití
- marc-ovskou strukturu doplňují:
 - *katalogizační pravidla* - jak zapisovat data do polí a podpolí
[AACR2: Anglo-American Cataloging Rules](#) , od 2015 [RDA: Resource Descript. and Access](#)
 - *soubory autorit* (národní, lokální)
 - *další instrukce a systémy*
(klasifikační pravidla, LCSH, ...)
- MARC pro **bibliografický záznam** x MARC pro **záznam autorit**

2.4 Proč soubory autorit?

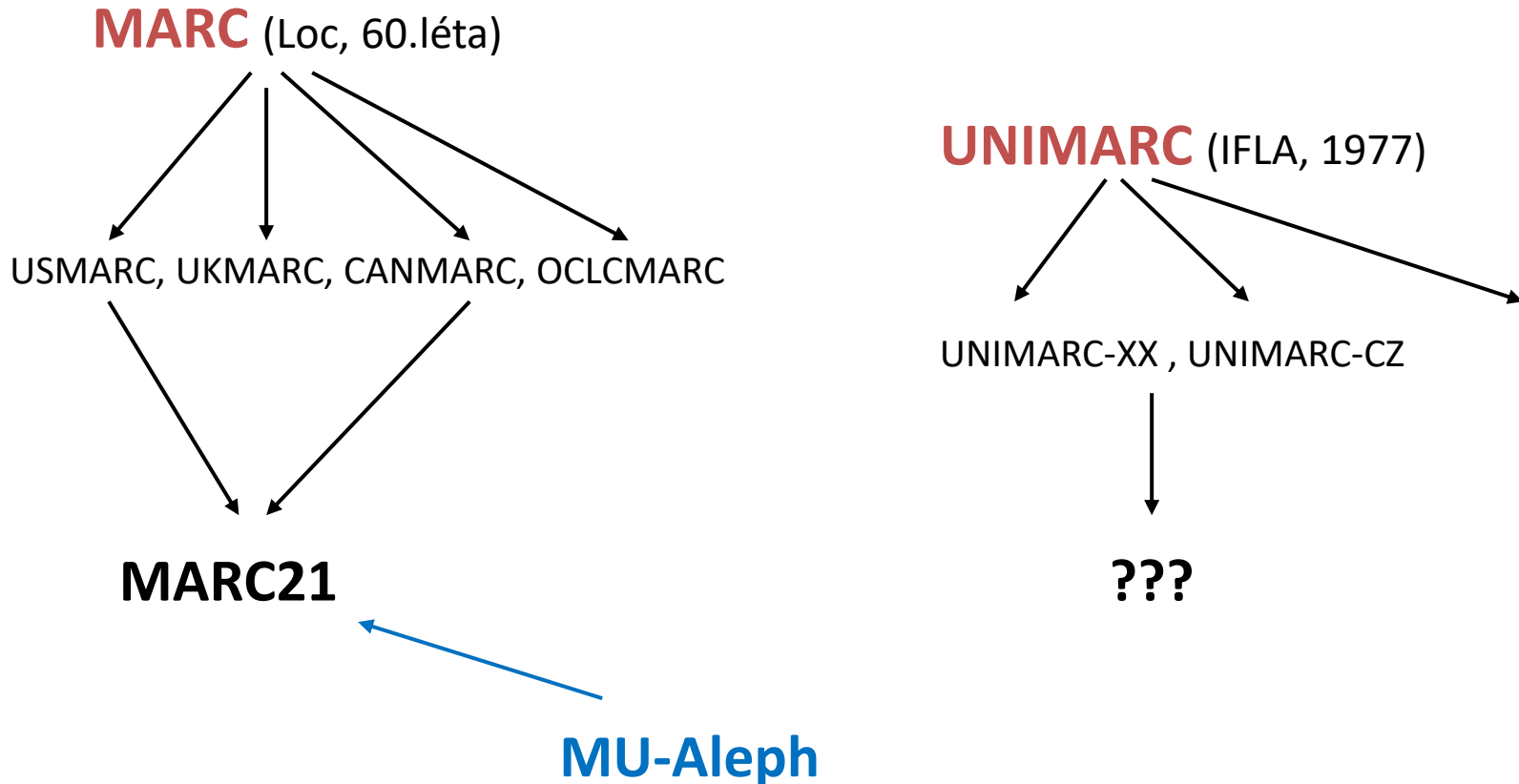
- 47 různých variant *jednoho jména* v dokumentech v LoC

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 Gaddafy
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

2.5 MARC a UNIMARC



<http://www.loc.gov/marc/>

<http://www.ifla.org/publications/unimarc-formats-and-related-documentation>

2.5 MARC a UNIMARC

- Odchylky UNIMARC x MARC21 (granularita, interpunkce, vazební pole)
UNIMARC: 700 \$aNNovák**\$bJ**an**\$f**1953- **MARC21: 100 \$a**Novák, Jan, **\$d**1953-
- UNIMARC = modernější
MARC21 = úspěšnější (podporovanější, LoC) – de-facto standard
- => **přechod na MARC21** (i v ČR - 2004)
- 2 způsoby zobrazení/uložení marcovské struktury:
 - **řádkový MACR** (přehlednější, pro člověka)
 - **výměnný formát** – ISO2709 (pro přenosy a strojové zpracování)
 - jedna dlouhá „nudle“ (formát pro MGP)

2.6.1 Řádkový UNIMARC

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

2.6.2 ISO 2709 UNIMARC

```
00718nam 2200217 45000010014000000005000900014010001600023
100004100039101000800080102000700088200014800095205001200243
210005400255215001000309225002500319675001300344711007200357
801003500429801002500464910001100489@CASLIN0000001@19960312@
$a80-7050-237-1@ $a19960305d1996 k y0czy0103 ba@0 $
acze@ $aCZ@1 $aZ'aznam pro souborný katalog$eUNIMARC$iTi's
t'ené monografie$fPracovní skupina CASLIN pro standardiza
ci a jmenné zpracování ... [et al.]@ $a1. vyd.@ $aPrah
a$cNárodní knihovna České republiky$d1996@ $a31 s.@1 $a
Standardizace$vč. 4@ $a025.3$9v@02$aCASLIN$bPracovní skup
ina pro standardizaci a jmenné zpracování@ 0$aCZ$bABA001
$c19960312$gAACR2$91@ 3$aCZ$bABA001$c19960515@ $aABA001@%
```

- **návěští** (24 úvodních znaků)
- **adresář** (12 znaků na každé pole): ident (3) + délka (4) + v.pozice (5)
- **tělo záznamu** (hodnoty polí včetně identifikace podpolí)

2.7 MARC – shrnutí

Plus: - velmi kvalitní, podrobná metadata
- MARCXML – revitalizace + oddělení datových prvků od struktury

Mínus: - cena 50-110 USD / úplný bibliografický záznam
- složitost tvorby (jen zaškolený katalogizátor)
- vhodnost pro digitální dokumenty ??

- LoC: 300.000 bibliografických záznamů ročně
- OCLC: 34 miliónů MARC-záznamů (30.000 člověkoroků práce)
- tradeoff: *cena metadat –versus- přínos pro vyhledávání*

- Digitální prostředí – změna ekonomického modelu:
 - přístup lze poskytnout i s malými investicemi do metadat
 - high-quality (drahá) metadata ospravedlnitelná pro velmi cenné zdroje

Síťová popisná metadata

3. RFC 1807



3.1 RFC 1807 – úvod

- první pragmatický pokus o jednoduchý Inet standard
- RFC 1807 : „**A Format for Bibliographic Records**“ (June 1995)
- Cornell – Dienst Protocol (NCSTRL)
- 29 prvků (4 povinné)
 - každý prvek začíná tagem ukončeným dvojicí dvojteček ::
 - na pořadí prvků nezáleží
 - prvek se může lib. opakovat (pořadí opakování je významné)
 - pouze ASCII znaky
- **keep it simple !**

dnes již spíš jen historická zajímavost

3.2 RFC 1807 – příklad

BIB-VERSION:: CS-TR-v2.1
ID:: OUKS//CS-TR-91-123
ENTRY:: January 15, 1992
ORGANIZATION:: Oceanview University, Kansas, Computer Science
TYPE:: Technical Report
REVISION:: January 5, 1995; FTP access information added
TITLE:: Scientific Communication must be timely
AUTHOR:: Finnegan, James A.
CONTACT:: Prof. J. A. Finnegan, CS Dept, Oceanview Univ, Oceanview, KS 54321 Tel: 913-456-7890
<Finnegan@cs.ouks.edu>
COPYRIGHT:: Copyright for the report (c) 1991, by J. A. Finnegan. All rights reserved. Permission is ...
HANDLE:: hdl:oceanview.electr/CS-TR-91-123
OTHER_ACCESS:: url:http://electr.oceanview.edu/CS-TR-91-123
OTHER_ACCESS:: url:ftp://electr.oceanview.edu/CS-TR-91-123
RETRIEVAL:: send email to Finnegan@cs.ouks.edu with fax number
KEYWORD:: Scientific Communication
CR-CATEGORY:: C.2.2 Computer Sys Org, Communication nets, Net Protocols
SERIES:: Communication
FUNDING:: FAS
CONTRACT:: FAS-91-C-1234
MONITORING:: FNBO
LANGUAGE:: English
NOTES:: This report is the full version of the paper with the same title in IEEE Trans ASSP Dec 1976
ABSTRACT:: Many alchemists in the country work on important fusion them....
END:: OUKS//CS-TR-91-123

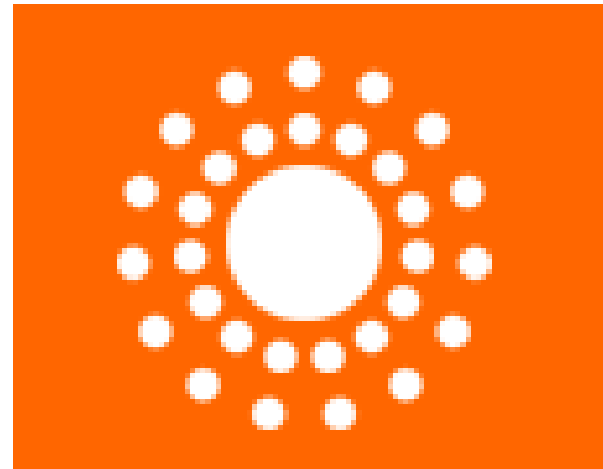
4. Dublin Core



4. DC – motivace

Motto:

- "The association of standardized descriptive metadata with networked objects has the potential for substantially improving resource discovery capabilities by enabling field-based (e.g., author, title) searches, permitting indexing of non-textual objects, and allowing access to the surrogate content that is distinct from access to the content of the resource itself." (Weibel and Lagoze, 1997)
- MARC = maximalistický přístup
- DC = minimalistický přístup





4.1 Dublin Core – požadavky

- x *klasická bibliografická metadata pro vyhledávání – příliš složitá*
- x *automaticky generovaná metadata vyhledávači – příliš jednoduchá*
- Dublin Core – **kompromis** : základní metadatové prvky (jádro) *společné pro popis jakéhokoliv zdroje*
- Hlavní požadavky
 - **jednoduchý** pro autorskou samokatalogizaci
 - **flexibilní** rozšiřitelný, pokud je třeba
 - **univerzální** co nejširší spektrum oborů a aplikací
 - **mezinárodní** mezinárodní konsensus, národní verze
 - **interoperabilita** základ pro sémantickou interoperabilitu
- nejčastější použití:
popis el.zdrojů za účelem jejich **efektivnějšího vyhledání na webu**

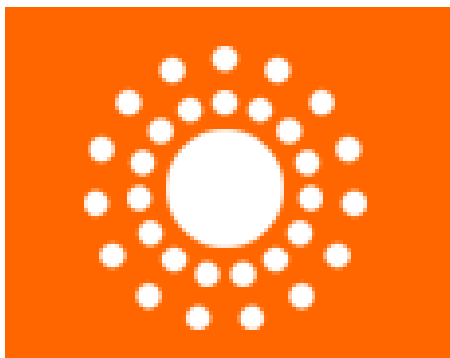


4.2 Vznik Dublin Core

- 1995 – ustavující workshop (Dublin-Ohio, OCLC/NCSA)
 - webaři
 - knihovníci
 - návrháři informačních systémů, ...
- **DCMI** (Dublin Core Metadata Initiative) – koordinace rozvoje
 - pracovní skupiny
 - každoroční workshop („valná hromada“)
- dvě verze DC:
 - **jednoduchý** („nekvalifikovaný“)
 - **kvalifikovaný**
- Jednoduchý DC: **15 základních prvků**



4.3 Jednoduchý DC



<i>obsah</i>	<i>vlastnictví</i>	<i>instance</i>
Title	Creator	Identifier
Subject	Contributor	Date
Description	Publisher	Language
Coverage	Rights	Format
Type		
Source		
Relation		
Audience	(Provenance, Rights Holder)	

- každý prvek nepovinný, opakovatelný, na pořadí prvků nezáleží
- pro každý prvek je dána jeho přesná sémantika
- syntaxi standard nedefinuje (záležitost doporučení, řízených slovníků)
- tvůrci aplikací mohou rozšířit základní sadu o své specifické prvky
- metadatový záznam buď vložen do zdroje nebo uložen externě
- popis prvků: http://www.ics.muni.cz/dublin_core/terms.html



4.3.1 Prvky DC

- **Title**
Def: Jméno dané zdroji.
Komentář: Nejčastěji bude jméno tvořit název, pod nímž je zdroj oficiálně znám.
- **Creator**
Def: Entita primárně zodpovědná za vytvoření obsahu zdroje.
Komentář: Příkladem může být osoba, organizace nebo služba.
- **Contributor**
Def: Entita, která přispěla k obsahu zdroje.
- **Date**
Def: Datum události, která se vyskytla během existence zdroje.
Komentář: Datum bude většinou spojeno s vytvořením nebo zpřístupněním zdroje.
- **Coverage**
Def: 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).



4.3.1 Prvky DC – příklad definice

- **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.



4.3.1 Prvky DC – příklad definice

- **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 Kvalifikovaný DC

- Jak zpřesnit popis zdroje pomocí DC ?
- Kvalifikovaný DC :
 - **kvalifikátor prvku**
(zužuje sémantiku: **Creator.Author, Creator.Illustrator**)
 - **kvalifikátor hodnoty**
(interpretace hodnoty: Datum=1994-04-12:**ISO8601**)
- *dumb-down* principle
- **DCMI Terms** = 15 základních prvků DC + 40 kvalifikátorů
<http://dublincore.org/documents/dcmi-terms/>
- Možnost definovat vlastní kvalifikátory (interoperabilita?)



4.4.1 Kvalifikátory

<i>prvek</i>	<i>kvalifikátor prvku</i>	<i>kvalifikátor hodnoty</i>
Title	Alternative	--
Subject	--	LCSH, MeSH, DDC, LCC, UDC
Description	Table of Contents Abstract	--
Date	Created, Valid, Issued	DCMI Period, W3C-DTF
Type	--	DCMI Type
Format	Extent, Medium	--
Identifier	--	URI
Source	--	URI
Language	--	ISO 639-2, RFC 1766
Relation	Is Version Of Has Version Is Replaced By, ...	URI
Coverage	Spatial Temporal	DCMI Point, ISO 3166, DCMI Box DCMI Period, W3C-DTF



4.5 Kvalifikovaný DC - příklad

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**

MYELEMENT.Checksum = **123456** : **XYZ**



4.5.1 Zápis 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>
```



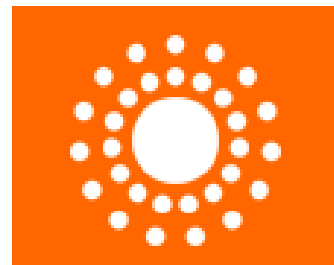

4.5.2 Zápis v XML

```
<?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://dublincore.org/2000/03/13/dcq#">
<rdf:Description rdf:about="http://dublincore.org/documents/2001/09/20/dcmes-xml/">
  <dc:title>Expressing Simple Dublin Core in RDF/XML</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>
  <dcq:isVersionOf rdf:resource="http://dublincore.org/documents/dcmes-xml/" />
  <dcq:replaces rdf:resource="http://dublincore.org/documents/2001/04/11/dcmes-xml/" />
</rdf:Description>
</rdf:RDF>
```

4.6 DC – aktuální stav

- **DCMI** (Dublin Core Metadata Initiative) – koordinace rozvoje
- Každoroční konference
[International Conference on Dublin Core and Metadata Applications](#)
- Internetovské doporučení – RFC 5013-2007
- de-facto standard přeložený do desítek jazyků
- český překlad http://www.ics.muni.cz/dublin_core
- standardizace: NISO Z39.85-2007
ISO 15836:2009
- využití DC v OAI
- nástroje pro generování DC-záznamů, ...
- <http://dublincore.org>

join us!





4.7 Zkušenosti z praxe

- "Developing a metadata scheme requires balancing richness and complexity against simplicity and ease of application. If it is too simple, the resulting description may not support desired functionality, but if it is too complex, few will apply it accurately or use it at all."

David Ruddy, DML 2010

- Zkušenosti z EuDML x DML-CZ (články)
 - název: podnázev? Více jazykových variant – originální název?
 - autor: pořadí autorů?
 - bibliografické údaje článku? (časopis, ročník, stránkování)
 - citované články?
 - dostupnost plného textu?
 - nadstandardní typy údajů?

tagovaný
nekvalifikovaný DC

```
<metadata>
  <dc:title>A note on on-line ranking number of graphs. 1. Simple graphs</dc:title>
  <dc:title>Poznámka k on-line pořadovým číslům grafů. 1. Jednoduché grafy</dc:title>
  <dc:creator>Semanišín, G.</dc:creator>
  <dc:creator>Soták, R.</dc:creator>
  <dc:date>2006</dc:date>
  <dc:language>eng</dc:language>
  <dc:publisher>Institute of Mathematics, Academy of Sciences of the Czech Republic</dc:publisher>
  <dc:publisher>Matematický ústav AV ČR</dc:publisher>

  <dc:relation>issn:0011-4642</dc:relation>
  <dc:relation>mr:MR2291759</dc:relation>
  <dc:relation>zbl:Zbl 1164.05360</dc:relation>
  <dc:relation>reference:[1] I.Brown: Extremal Graph Theory. Academic Press, London, 1978.</dc:relation>

  <dc:identifier>http://hdl.handle.net/10338.dmlcz/128089</dc:identifier>
  <dc:identifier>bibliographicCitation:Czechoslovak Mathematical Journal, vol. 56 (2006), issue 2, pp. 591-599</dc:identifier>
  <dc:identifier>bibliographicCitationStructured:JournalTitleFull:Czechoslovak Mathematical Journal; JournalVolume:56; JournalIssueDate:2006; JournalIssueNumber:2; JournalPages:591-599</dc:identifier>

  <dc:format>application/pdf</dc:format>
  <dc:type>type:math</dc:type>
  <dc:type>text:journal_article</dc:type>

  <dc:subject>keyword:on-line ranking number</dc:subject>
  <dc:subject>msc:05C15</dc:subject>
  <dc:description>summary:A  $k$ -ranking of a graph  $G=(V,E)$  is a mapping such that each path with endvertices of the same colour  $c$  contains an internal vertex...</dc:description>

  <dc:rights>access:Unrestricted</dc:rights>
  <dc:rights>rights:DML-CZ Czech Digital Mathematics Library, http://dml.cz/</dc:rights>
  <dc:rights>conditionOfUse:http://dml.cz/use</dc:rights>
```

pořadí!



4.8 DC – zhodnocení

Dublin Core

- **velký úspěch**
 - rozšíření
 - DC – stavební kámen pro další technologie (OAI-PMH)
- **nenaplněná očekávání**
 - rozpor mezi jednoduchostí a použitelností
 - pro řadu aplikací DC nepoužitelný
- **Nový favorit = MODS ?**

5. Metadatové schéma LC



5. LoC – Core Metadata Elements



- Historický „firemní standard“ pro *National DL Program*, LoC
- *strukturální, administrativní*, částečně i *popisná* metadata pro podporu všech funkcí DL (řízení přístupu, vyhledávání, prezentace, administrace, identifikace, archivace)
- **5-stupňová úroveň metadat:**
 - **set** jedna digitální sbírka; tvoří ji jeden či více *agregátů*
 - **agregát** skupina DO stejného typu (text,video) či správy
 - **primární objekt** ekvivalent fyzické položky - kniha, zvuková nahrávka
 - **meziobjekt** různý pohled, komponenta primárního objektu (kniha: 1. naskenované obrázky jednotlivých stran, 2. prohledávatelné texty stran)
 - **terminální objekt** soubor/y s digitální informací
- **77 metadat. prvků** <http://lcweb.loc.gov/standards/metadata.html>

5.1 Příklad

- Fyzická sbírka [Alexander Graham Bell Family Papers](#)
- Digitální sbírka [AM, The A.G.Bell Family Papers, 1862-1939](#)
<http://memory.loc.gov/ammem/bellhtml/bellhome.html>
- Artefakt [Dopis od A.M.Bella → A.G.Bellovi z 2.3.1864](#)
- Popis [jeden list oboustranně psaného rukopisu, plus přepis na psacím stroji \(2 stránky\)](#)

-
- celkem 4 různé obrázky plus textový přepis
 - náhledová stránka obsahuje náhled jednoho obrázku (96 dpi), navigační odkazy na další/předchozí navigační stránku, odkaz na referenční obrázek v plném rozlišení, na domovskou stránku, ...
 - referenční archivní obrázky jsou v rozlišení 300 dpi
 - metadata pro tento 1 artefakt tvoří 16 metadatových záznamů

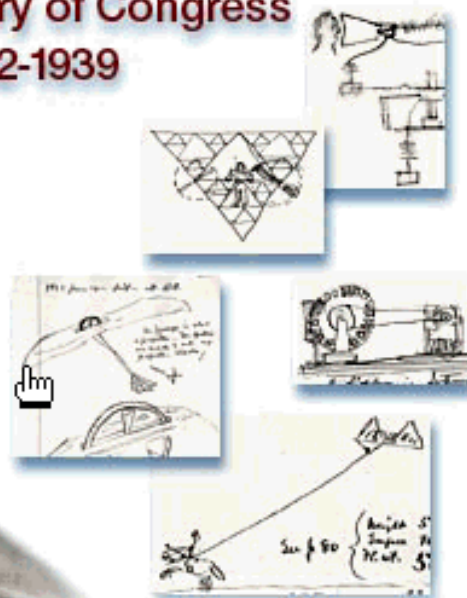
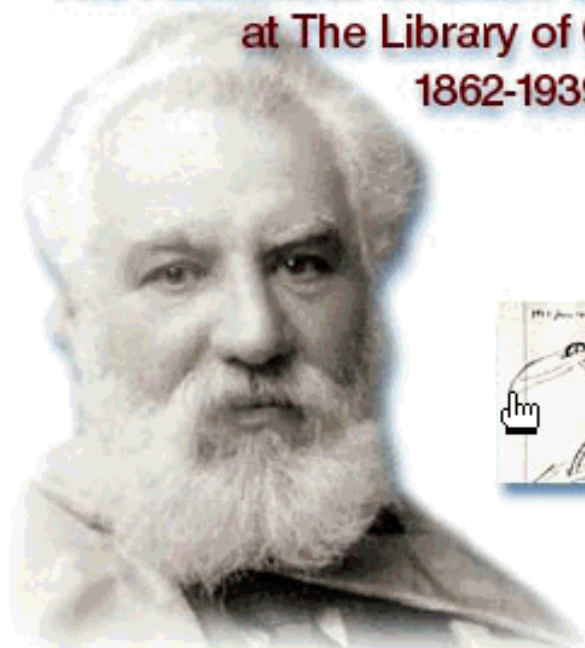
5.2 A.G.Bell Family Papers

AMERICAN MEMORY



LIBRARY of CONGRESS

The Alexander Graham Bell Family Papers at The Library of Congress 1862-1939



[Manuscript Division, Library of Congress](#)

Search by [Keyword](#) | Browse by [Series](#) | [Subject](#) | [Name](#)

The online version of the Alexander Graham Bell Family Papers at the Library of Congress will comprise a

5.2 A.G.Bell Family Papers

The Alexander Graham Bell Family Papers

Letter from Alexander Melville Bell to Alexander Graham Bell, March 2, 1864 - [Transcription](#)
(Series: Family Papers, Folder: Alexander Melville Bell, Family Correspondence, Alexander Graham Bell, 1862-1866)

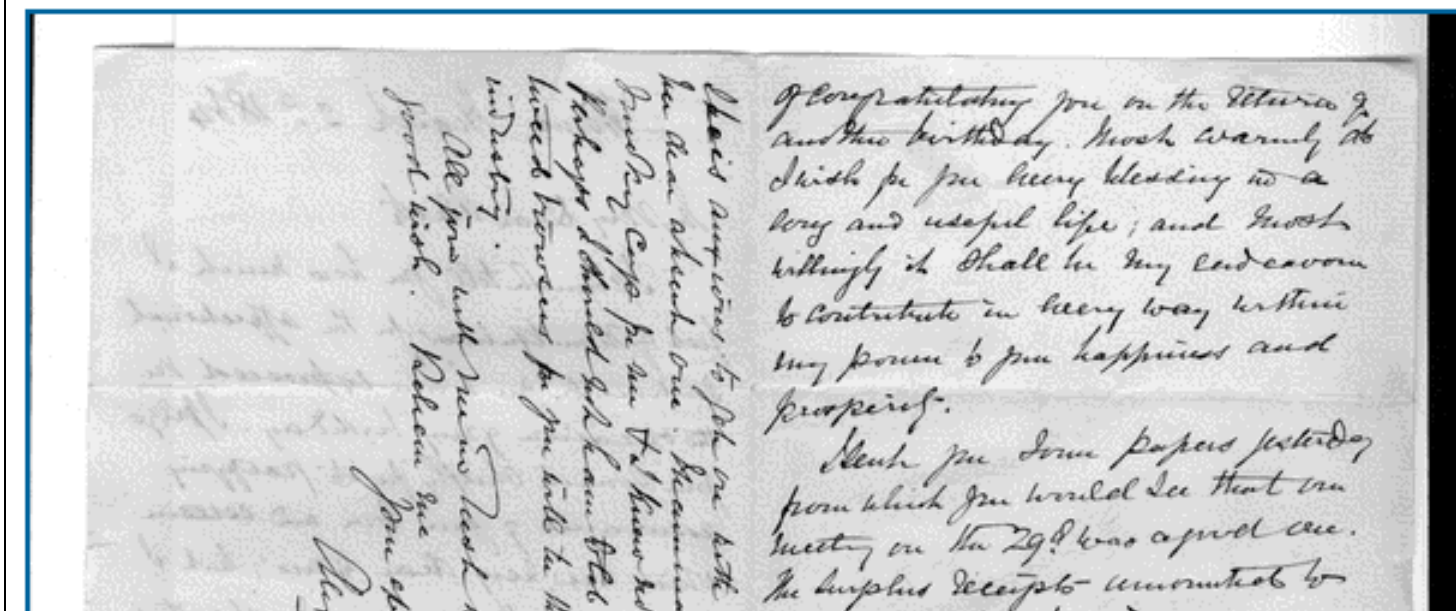
Image 2 of 4

Turn to image

2

[PREV IMAGE](#) | [NEXT IMAGE](#)

[Bell Papers Home](#) | [Archival grayscale/color \(JPEG - 343K\)](#)



5.2 A.G.Bell Family Papers

The Alexander Graham Bell Family Papers

Letter from Alexander Melville Bell to Alexander Graham Bell, March 2, 1864 - [Transcription](#)
(Series: Family Papers, Folder: Alexander Melville Bell, Family Correspondence, Alexander Graham Bell, 1862-1866)

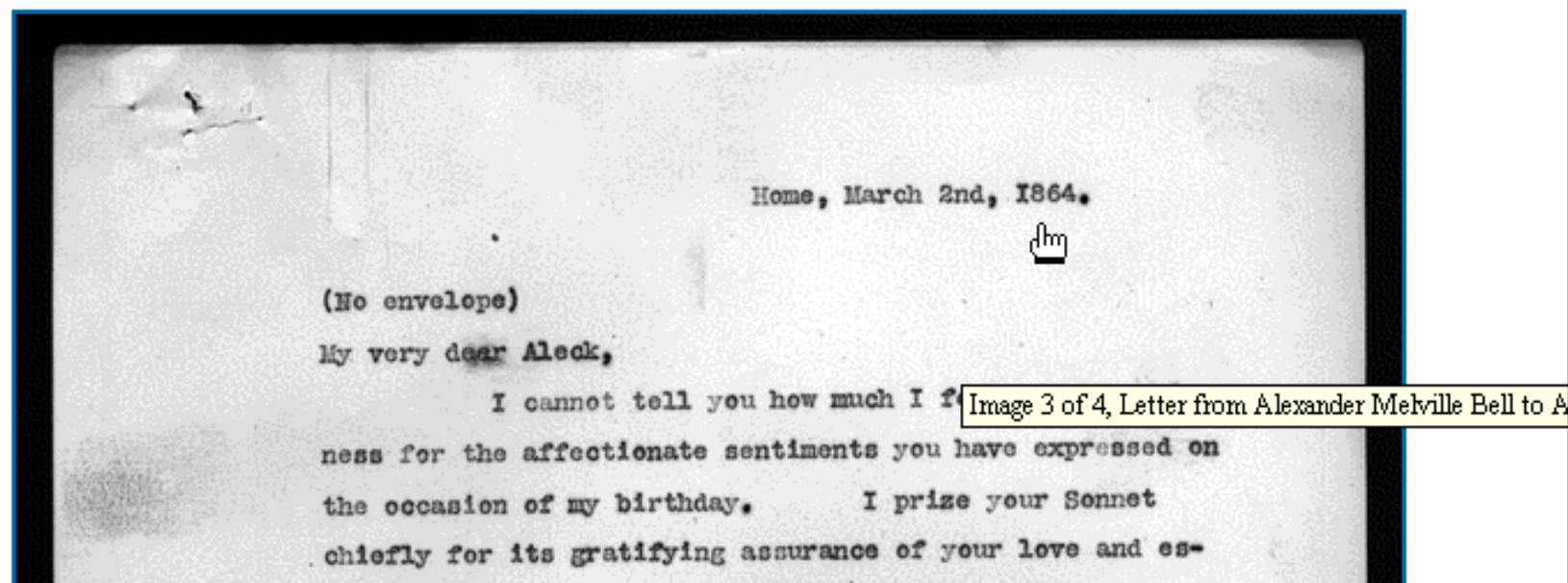
Image 3 of 4

Turn to image

3

[PREV IMAGE](#) | [NEXT IMAGE](#)

[Bell Papers Home](#) | [Archival grayscale/color \(JPEG - 451K\)](#)



5.2 A.G.Bell Family Papers

struktura metadatových záznamů pro dopis A.M.Bella

- **Set** **Alexander Graham Bell Family Papers**
- **Agregate** **1 záznam: jediný typ dokumentů**

• PrimaryO	dopis AMB z 2.3.1864		
• Meziobjekt	SGML verze	náhled.stránky	ref.arch.obrázky
• TerminalO	SGML soubor	1.str rukopisu	1.str rukopisu
	linky na obr.	2.str rukopisu	2.str rukopisu
	linky na SGML syst.soub	1.str strojopisu	1.str strojopisu
	podpora pro SGML browser	2.str strojopisu	2.str strojopisu

- **16 metadatových záznamů pro 1 objekt sbírky !**

Literatura



Doplňková literatura

- NISO 2017: [Understanding metadata](https://www.niso.org/publications/understanding-metadata-2017).
<https://www.niso.org/publications/understanding-metadata-2017>
- [Dublin Core Metadata Initiative – Metadata Basics](http://dublincore.org/metadata-basics/)
<http://dublincore.org/metadata-basics/>