

Digitální knihovny

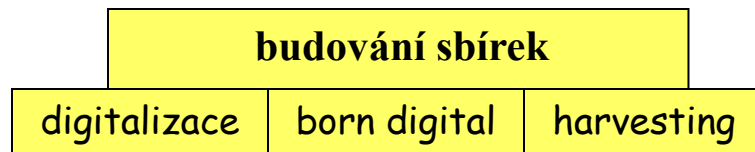
Identifikátory



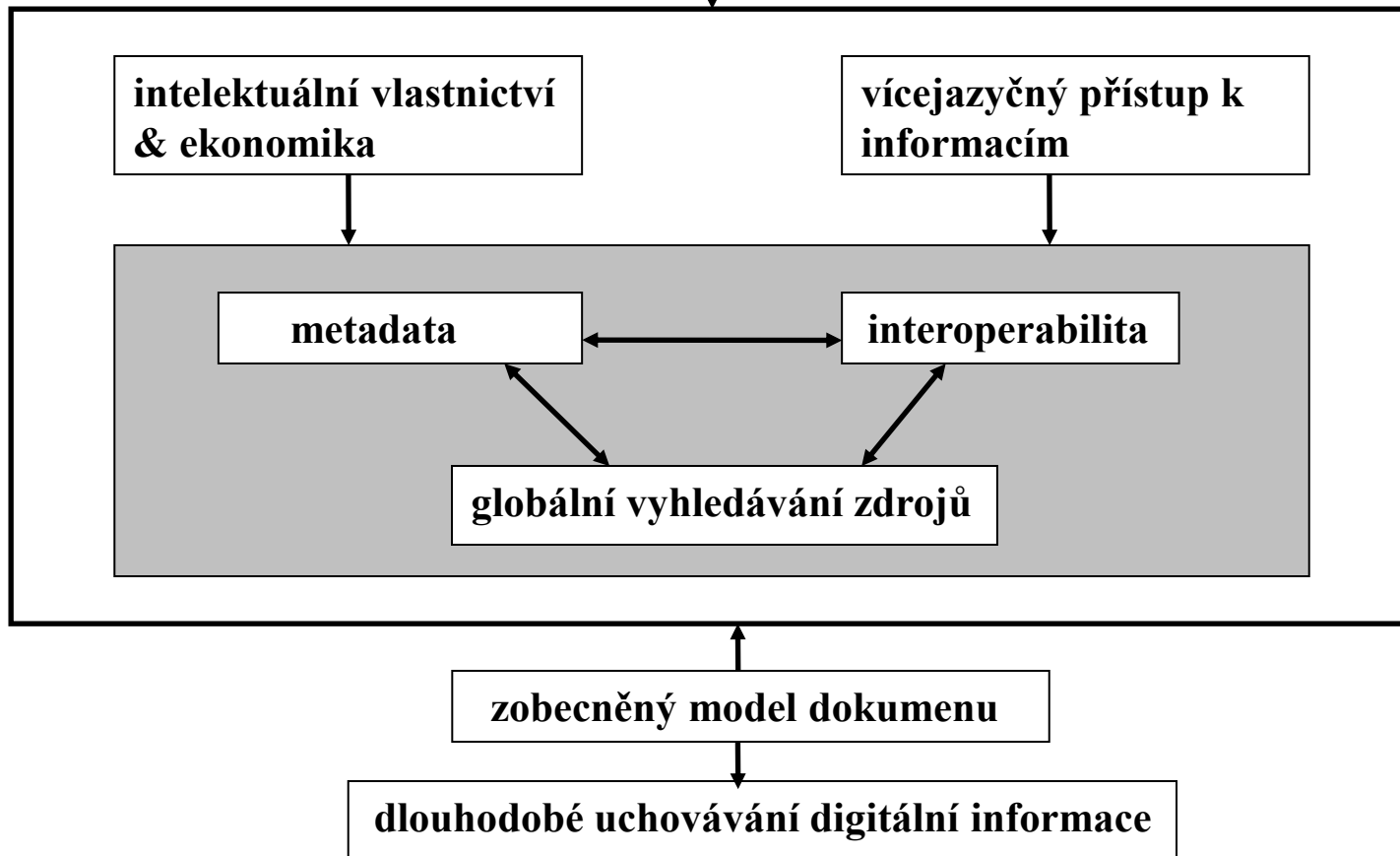
Miroslav Bartošek

Ústav výpočetní techniky MU

Knihovnicko-informační centrum MU



obecný rámec a architektura DL



Obsah přednášky

Hierarchická abstrakce intelektuálního díla

1. O identifikátorech obecně
2. Klasické (knihovnické) identifikátory
(ISBN, ISSN, SICI, BICI, ISTC, ISNI)

”digitální” identifikátory

3. URN
4. PURL
5. Handles
6. DOI
7. (ARK)

Hierarchická abstrakce intelektuálního díla (IFLA model)



Abstrakce intelekt-díla

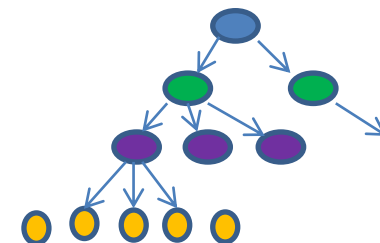
- Uživatelé chtějí odkazovat na intelektuální výtvořy na vyšší úrovni abstrakce než je počítačový soubor nebo konkrétní provedení výtvořu
- Potřeba přesnější specifikace intelektuálního výtvořu a jeho forem
 - „Napište eseř o [Babičce od Boženy Němcové](#)“
(míněno dílo samo o sobě, ne nějaké jeho nějaké konkrétní vyřádření či provedení)
 - „Jdeš do knihkupectví? Kup mi [Babičku od Boženy Němcové](#)“
(míněn vřtisk konkrétního knižního vydání díla)
 - Upomínka z knihovny:
„Vraťte knihu [Babička od Boženy Němcové](#) vypůjčenou dne 14.10.2015“
(míněna konkrétní jednotka od konkrétního knižního vydání)
- Lidé jsou schopni porozumět, o co jde – podle kontextu
Stroje to potřebují specifikovat přesně, polopatiřticky

Abstrakce intelektuálního díla

- **IFLA model 1998** – kategorizace intelektuálních děl dle různé úrovně abstrakce (4 úrovně):
 - *work (dílo)*
 - *expression (vyjádření)*
 - *manifestation (projev)*
 - *item (jednotka)*
- *Součástí specifikace [Functional Requirements for Bibliographic Records](http://www.ifla.org/VII/s13/frbr/frbr.htm) IFLA, 1998.* <http://www.ifla.org/VII/s13/frbr/frbr.htm>
IFLA = International Federation of Library Associations
- využití v mnoha oblastech (DL, knihovny, e-commerce, ...)
- možnost strojového propojení dílo – vyjádření – projevy – jednotky

IFLA model FRBR

- **work (dílo)** *obecná abstraktní idea*
intelektuální/umělecký výtvar jako obecný *abstraktní pojem* nejvyšší úrovně (Homérova Illiada, Beethovena 5.symfonie, operační systém Unix, píseň *Let it be* od Beatles)
- **expression (vyjádření)** *vyjádření ideje*
intelektuální či umělecká realizace, časoprostorová fixace díla v nějaké formě (Illiada napřed ústně, pak zápisem od XY; symfonie jako originální partitura autora nebo konkrétní hudební nahrávka od ABC; text písně *Let it be* v anglickém originálu, český překlad textu od XY)
- **manifestation (projev)** *prezentace vyjádření v určitém formátu*
konkrétní fyzické „*zhmotnění*“ nějakého vyjádření (přepisy textu Illiady v konkrétních knižních vydáních; nahrávka symfonie od ABC vydaná jako zvukový záznam na CD, mag. kazetě či video-záznam televizního vysílání)
- **item (jednotka)** *jedna konkrétní jednotka daného projevu*
kopie jednoho projevu díla (výtisky dané knihy, kopie souboru)
- Některé druhy intelektuálních děl lze obtížně „napasovat“ [např. namalovaný obraz: jednotka je totožná s dílem];



dílo
vyjádření
projev
jednotka

IFLA model – příklad

- ***work (dílo)***
„Babička“ od B. Němcové – jako dílo jako takové (bez ohledu na jeho vyjádření a provedení)
- ***expression (vyjádření)***
rukopis od Němcové z roku 1854; redakční úprava rukopisu BN od redaktora XY z roku 1855; upravený/modernizovaný rukopis BN od editora UV z roku 2003; anglický překlad vydání z roku 2003 od překladatele G.W.Smith z roku 2008; scénář filmu Babička od Františka Čápa z roku 1940; film Babička režiséra Františka Čápa z roku 1940 s Terezií Brzkovou v hlavní roli; film Babička režiséra Antonína Moskalyka z roku 1971 s Jarmilou Kurandovou; rozhlasová adaptace režiséra XY z roku 1965; ...
- ***manifestation (projev)***
kniha vydaná nakladatelem Jaroslavem Pospíšilem v roce 1855 (podle redakčního textu XY z roku 1855); páté vydání knihy z roku 1900 (podle původního redakčního textu) s ilustracemi Adolfa Kašpara; originální kopie filmu z roku 1940 na celuloidovém filmovém pásu; tentýž film přepsaný na DVD a vydaný v roce 2005 v edici MF Zlatý fond; ...
- ***item (jednotka)***
výtisky vydání Babičky z roku 1900; ...

1. O identifikátorech



1.1 Proč identifikátory

- Eliminace fyzického kontaktu znamená vyšší potřebu identifikace (e-commerce)
 - přesnost
 - spolehlivost
 - strojová zpracovatelnost + propojování informací
- **Identifikátor = jméno jednoznačně identifikující objekt v dané doméně**
 - Příklady: rodné číslo, (obyvatel ČR) UČO (student MU), IČ/DIČ, SPZ, ISBN, ...
- identifikátory lokální x globální
- „If there is one thing that distinguishes a digital library from a mere web site, it is that libraries do their best to provide *reliable, persistent access through durable links* [identifiers].“
(J.A.Kunze, California Digital Library)

1.1 Historická vsuvka: pojmenovávání osob

- Historický vývoj **pojmenovávání osob** – příklad rostoucích požadavků na přesnější identifikaci v souvislosti s rostoucí mobilitou (rozšiřováním domény)
 - Původně jen osobní (křestní) jméno (Janek) – postačující v malé lokální komunitě
 - S růstem velikosti sídel a vyšší mobilitou – potřeba přesnější identifikace
 - **Přídomek** u šlechty – (Albrecht z Valdštejna) – IT, FR **od 10. století**
 - **Příjímí** – (přízvisko: Matěj Janovec (Janův), Matěj řezník) – neoficiální nedědičné druhé doplňkové jméno, nejprve u městského obyvatelstva – sporadicky **od 14. st.**, hojněji od 16.st.
 - **Povinné používání příjmení** – patent Josefa II. z roku **1786**
 - **Číslo občanského průkazu** – od 1948
 - **Rodné číslo** v Československu – **od 1953** (oddělení čísla průkazu od čísla občana)
- kuriozitka: **Četnost příjmení nebo jména v ČR** - <https://www.kdejsme.cz/>
 - Rozšíření příjmení/jména v rámci ČR
 - 412 tisíc unikátních příjmení / 148 tisíc jmen
 - Žebříček nejčastějších příjmení/jmen (kvíz na „top-10“)



Příjmení

Jméno

Bartošek

Vyhledat

Hustota

Počet

Zadejte příjmení nebo jméno zjistíte jeho četnost

Bartošek (865)

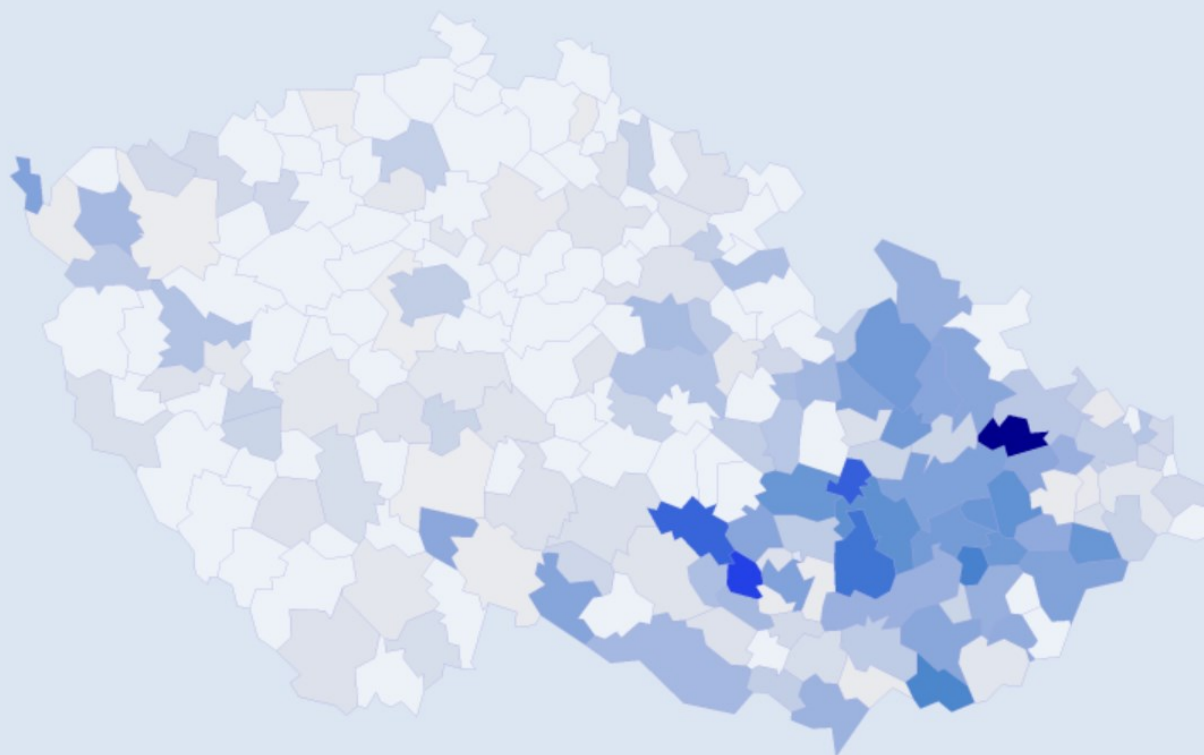
Obec s RP	Počet	Každý X-tý
Praha	71	19038
Brno	64	7355
Prostějov	36	4597
Olomouc	32	7163
Vyškov	30	3467
Přerov	25	6014
Šumperk	24	5760
Velké Meziříčí	22	3257
Veselí nad Moravou	21	3778
Ostrava	21	19497
Boskovice	20	5122
Uherské Hradiště	18	8736
Vsetín	18	7421
Rosice	17	2874
Zlín	16	10411
Pardubice	15	12802
Vítkov	15	1886
Hranice	15	4667
Rožnov pod Radhoštěm	14	5125
Znojmo	13	12123
Kroměříž	13	10494
Sokolov	12	12184
Holešov	12	3663
Břeclav	11	11020
Chrudim	10	14923
Opava	10	16872

[více ...](#)

Reklamu zavřela
společnost
Google

Příjmení: 'Bartošek', počet výskytů 865 v celé ČR

Hustota: 1 'Bartošek' na 12371 obyvatel ČR, průměrný věk: 41



© KdeJsme.cz

Historický vývoj četnosti příjmení "Bartošek":

- 2010: Počet výskytů **příjmení "Bartošek"** v roce 2010: 884 v celé České republice

[Odkud data čerpáme?](#)

1.2 Charakteristiky a vlastnosti ID

- 1. forma**
- 2. jednoznačnost (globální)**
- 3. persistence**
- 4. směrovatelnost**

- rozsah použitelnosti identifikátoru (pro jaké objekty)
- metadata pro popis objektu spojená s identifikátorem

1.2.1 Forma ID

- **inteligentní** (structured)
id v sobě nese určitou informaci, má sémantiku
- „**hloupý**“ (dumb, opaque)
ze samotného identifikátoru nelze nic odvodit
- **vypočítatelný**
id lze spočítat z obsahu/vlastností objektu

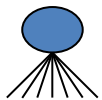
Trend k „hloupým“ identifikátorům – lepší persistence, bezpečnost, ochrana osobních dat, ...

(nejen v digitálním světě – viz SPZ)

1.2.2 Jednoznačnost ID

Jak zajistit globální jednoznačnost identifikátorů ?

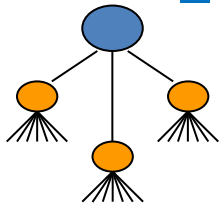
Způsob přidělování identifikátorů



– **centralizovaný**

ISSN

1 centrum přiděluje ID všem, a hlídá jednoznačnost



– **distribuovaný**

ISBN, RČ

oprávnění přidělovat ID distribuováno lokálními autoritám

- 1 globální centrum přiděluje globálně-jednoznačné ID lokálním centrům
- lokální centra přidělují lokálně-jednoznačné ID
- globální jméno: **ID_lokálního_centra/lokální_ID** (prefix/sufix)

Distribuované přidělování pružnější, efektivnější, škálovatelné...
...(ale implementačně náročnější)

1.2.3 Persistence ID

- persistence = **trvanlivost** (trvalá platnost identifikátoru)
- jednou přidělený identifikátor trvá/platí/funguje „na věky“ (i po případném zániku organizace, která jej přidělila)
- persistence není vrozenou vlastností žádného identifikátoru (identifikačního schématu) sama o sobě, **vždy je to záležitost služby - společenské, organizační a technické infrastruktury**
 - Rodné číslo: při změně pohlaví se osobě přiděluje nové RČ (persistence??)
- Problémy s URL:
 - s persistencí („poločas rozpadu“ identifikátorů URL)
 - identifikuje LOKACI, nikoliv objekt

1.2.4 Směřovatelnost ID

- směrovatelnost – resoluční služba, která umí na síti **automaticky (on-line) propojit daný identifikátor s objektem**, který je tímto identifikátorem označený
- laicky:
zadám identifikátor do prohlížeče a dostanu příslušný objekt (či jeho záznam) – ať je uložen v Internetu kdekoliv (viz URL)
- **akční identifikátory**
- směrovatelnost URL (= resoluce přes DNS)

1.2.5 Hierarchický systém ID

Neexistuje 1 univerzální ID pro vše => potřeba vícevrstvé identifikace

- **organizace (knihovna)**

 - ISIL** Intl Standard Identifier for Libraries and Related Organizations

- **sbírka, služba**

 - ISCI** Intl Standard Collection Identifier

- **autor/korporace**

 - ISNI** Intl Standard Name Identifier

- **dílo**

 - ISTC** Intl Standard Text Code

 - ISWC** Intl Standard Musical Work Code

- **projev**

 - ISBN** Intl Standard Book Number

 - ISSN** Intl Standard Serial Number

 - ISMN** Intl Standard Music Number

- **komponenta**

 - SICI** Serial Item and Contribution Identifier

 - DOI** Digital Object Identifier

Identifikátory jako
mezinárodní standardy

1.3 Kdo vytváří standardy?

- **mezinárodní (de-jure)**

- **ISO – International Organization for Standardization**

- TC 46 – **Technical Committee** for information and documentation standards

- SC 9 – **Subcommittee** on identification and description of information resources

- http://www.iso.org/iso/iso_technical_committee.html?commid=48836

- **IETF – Internet Engineering Task Force**

- **W3C – World Wide Web Consortium**



- ISO xxxx

- RFC xxx

- XML, RDF

- **národní**

- **ANSI – American National Standards Institute**

- NISO – National Information Standards Organizations

- Z39.56

- **ÚNMZ – Úřad pro technickou normalizaci, metrologii a státní zkušebnictví**

- **oborový (de-facto)**

- Kongresová knihovna

- MARC

- Asociace amerických nakladatelů

- DOI

- Dublin Core Metadata Initiative

- DC

2. Klasické identifikátory



2. Klasické „knihovnické“ ID

ISBN	ISO 2108:2005	International Standard Book Numbering (ISBN) 1972
ISSN	ISO 3297:2007	International Standard Serial Number (ISSN) 1975
ISRC	ISO 3901:2001	International Standard Recording Code (ISRC) 1986
ISRN	ISO 10444:1997	Intl Standard Technical Report Number (ISRN) ke zrušení?
ISMN	ISO 10957:2009	International Standard Music Number (ISMN) 1993
ISWC	ISO 15707:2001	International Standard Musical Work Code (ISWC) 2001
ISAN	ISO 15706:2002	International Standard Audiovisual Number (ISAN) 2002
V-ISAN	ISO 15706-2:2007	Version Identifier for Audiovisual works (V-ISAN) 2007

...

<https://www.iso.org/committee/48836/x/catalogue/>

2.1 ISBN



- **ISBN 80-00-01987-6** **International Standard Book Number**
- „inteligentní“ identifikátor, fixní délka (10 znaků)
- ISO norma od 1972, v ČSSR v provozu od 1989
- Mezinárodní agentura ISBN (Berlín), národní agentury (NK ČR)
- struktura – 4 části proměnlivé délky:
 1. země nakladatele (1-anglická oblast, 80 ČR a SR; přiděluje MA)
 2. nakladatel (2-7 cifer dle velikosti produkce; přiděluje národní agentura)
 3. konkrétní vydání v rámci nakladatele (až 6 cifer; přiděluje nakladatel)
 4. kontrolní znak (1-9,X)
- ISBN 8 0 0 0 0 1 9 8 7 suma: **159**/11=14, zb.5
 váhy 10 9 8 7 6 5 4 3 2 *check*: 11-5 = **6** (**10→X**)
 součiny **80 0 0 0 0 5 36 24 14** *kontrola*: *mod*(159+6,11)=0
- <http://www.isbn-international.org/>

2.1 ISBN a web



- Jasná pravidla pro přidělování ISBN u klasických tištěných publikací...
 - ... ale problémy s aplikací ISBN v digitálním světě
 - web : nakladatelem kdokoliv (exponenciální nárůst nakladatelských čísel)
 - dramatický nárůst počtu e-publikací (plus přidělování ISBN pro kapitoly, ...)
- => [krize 1998: prostor čísel ISBN se brzy vyčerpá !!!](#)
- **ISBN-13** (od 2007)
 - kompatibilita se systémem EAN (“evropský” obchodní kód zboží)
 - **přidat před ISBN prefix 978 (knihy) a 979 (hudebniny) = 2x větší prostor**
 - ISBN **978-316148410-0**
 - ISBN-A (actionable) – ISBN registrované jako identifikátor DOI
 - Pracovní skupina v rámci TC46/SC9 pro přípravu nového standardu ISBN

2.2 ISSN



- **ISSN 0167-8191** **International Standard Serial Identifier**
- pro seriálové publikace, včetně elektronických (tituly časopisů, ...)
- „hloupý“ identifikátor, fixní délka 8 cifer
(poslední cifrou je kontrolní znak, algoritmus jako u ISBN)
- ISO norma od 1975
- centrální přidělování a správa (Paříž, Mezinárodní centrum pro ISSN)
národní střediska ISSN – prostředníci pro komunikaci s centrem, národní metadata)
- 1 milión záznamů v **centrálním registru ISSN** (kapacita 10 mil)
pro každé přidělené ISSN existuje v DB metadatový záznam o daném seriálu
- el.časopisy zatím kapacitu neohrožují, ale:
 - krátký poločas rozpadu
 - nová katalogizační pravidla (ISSN pro každou www-stránku s dokumenty)

<http://www.issn.org>

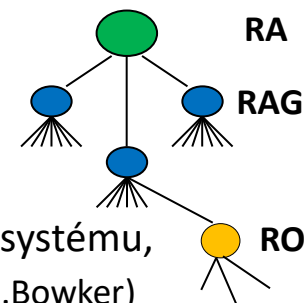
2.3 SICI – články v seriálech

- **0730-9295(199206)11:2<168:CRFAOC>2.0.TX;2-#**
 - M.Needleman. Computing Resources for an on-line catalog – 10 years later. *Information technology and libraries*. 11(2), červen 1992, pp. 168-175
- „strojově-generovatelný“ identifikátor – **Serial Item Component Identifier**
 - ISSN
 - item segment (údaje o čísle seriálu)
 - contribution segment (údaje o článku)
 - control segment (verze standardu 2.0, typ TX - „tištěný text“)
- celosvětově se neprosadil (pouze US ANSI standard, od 1996: Z39.56 – 1996 (R2002))
- **BICI – Book Item Component ID** podobné principy (nedokončený US ANSI/NISO draft)
 - identifikace samostatně „prodejné“ jednotky (kapitoly, obr, poznámky, ...) knihy s přiděleným ISBN
 - nedokončený US ANSI/NISO draft – „dead on arrival“

2.4 ISTC – textová díla

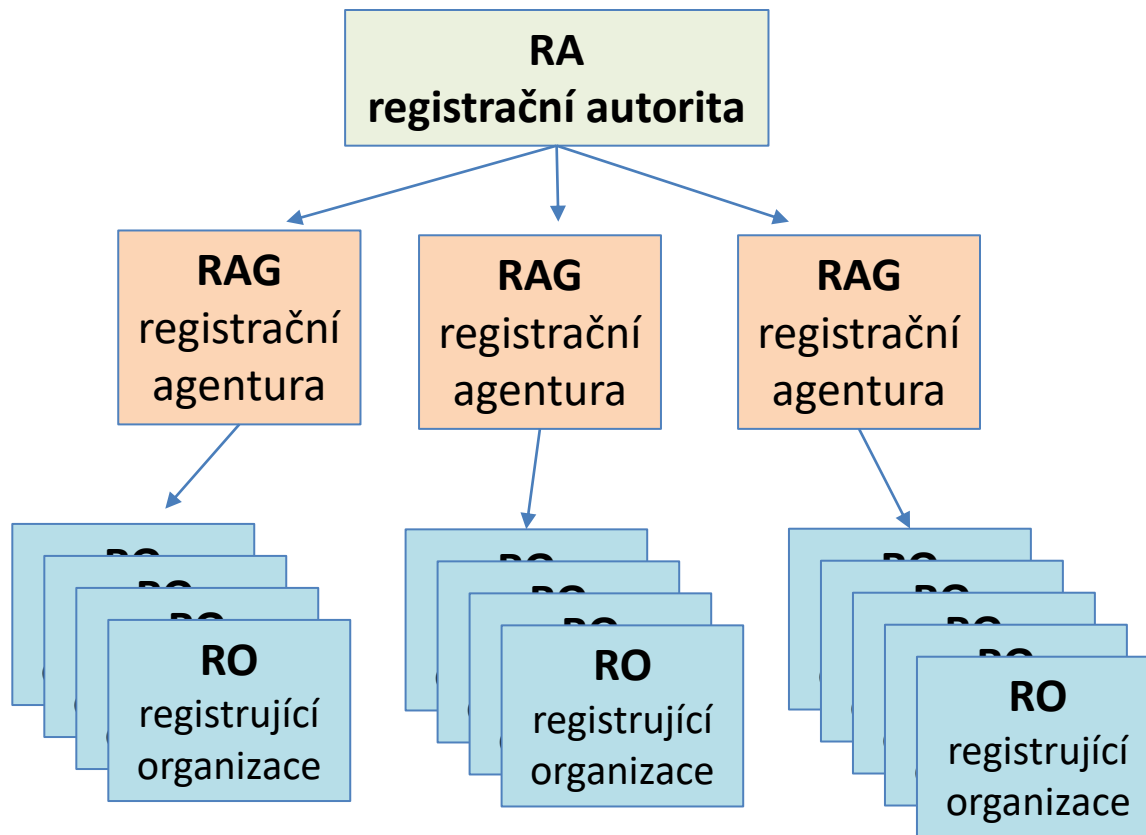
- **ISTC 0A9-2002-12B4A105-6** **International Standard Text Code**
- identifikátor fixní délky (16 znaků) – 4 části: (Identifikátor díla – viz IFLA-FRBR)

- kód registrační agentury (až 4096)
- rok
- identifikátor intelektuálního díla (ne projevu! až miliarda/rok)
- kontrolní číslice



- **RA – registrační autorita:** International ISTC Agency (2008) = administrace systému, kódy RAGs, centrální databáze ISTC záznamů (konsorcium CISAC/Nielsen BookData/R.R.Bowker)
- **RAGs – registrační agentury:** služby koncovým uživatelům, přidělování ISTC, metazáznamy, interface uživatelů na RA (2017/10 – již 8 RAGs)
- **RO: registrující organizace:** autoři, agenti, vydavatelé, knihovny, kolektivní správci práv
- Katalogizace všech intelekt. děl (monografie, články, ...) = obrovský úkol, kooperace
- Co je „textové“ dílo? (knihy, články, eseje, básně, hry, ..., ale i jejich části – monolog z hry, sloka z básně, ... Jde o díla dle IFLA klasifikace vyjádřená výhradně/hlavně v textové formě)
- <http://www.istc-international.org/>
☹ Zatím se příliš neujal, probíhá revize ISO standardu... (dobrý úmysl nestačí)

Organizační struktura pro nové identifikační standardy ISO



Administrace a rozvoj systému

Služby zákazníkům v určité oblasti nasazení daného id-systému

Přidělování identifikátorů koncovým objektům a využívání systému v dané organizaci

2.5 ISNI – osoby a korporace

Přehršel „globálních“ identifikátorů osob (ale teprve v posledních letech)

- **ISNI** – International Standard Name Identifier
 - ISO standard od 2012, aktuálně 11,5 mil přidělených ID, perspektivní? Pomalý náběh...
- **VIAF** – Virtual International Authority File
 - společný projekt národních knihoven od 1998, propojení národních autoritních-záznamů
- **ORCID** – Open Researcher and Contributor
 - od 2012, komunitní projekt, přes 65 členských institucí vč. velkých vydavatelů – Nature, Elsevier, Springer, Thomson-Reuters aj.; propojení na ISNI (subset ISNI)
- **ResearchID**
 - Identifikátor výzkumníků v databázích Web of Knowledge, Thomson-Reuters; propojení na ORCID
- **ScopusID**
 - Scopus Author Identifier (Elsevier)
- **OpenID a mojID**
 - od 2007, otevřený standard pro decentralizovanou autentizaci uživatelů na webu (OpenID = URL) – AOL, BBC, Google, IBM, MySpace, PayPal, seznam.cz ... (jeden účet pro přihlašování k webovým službám více poskytovatelů); v ČR: mojID – CZ.NIC

2.5 ISNI – osoby a korporace

ISNI 0000 0000 7988 7687 (Bartošek, Miroslav) ISO od 2012

- 16 cifer, poslední kontrolní znak (0-9,X) – 100.000 miliard kombinací
- identifikace fyzických osob, právnických osob, fiktivních postav
- propojení jmenných identit napříč různorodými systémy
- aktuálně 11,5 mil ID (z toho 10,5 pro fyzické osoby, 1 pro organizace)
- <http://www.isni.org/do-you-have-an-isni>

- **RA – registrační autorita:** ISNI International Agency – konsorcium velkých vydavatelů, knihoven aj. - rozvoj a financování systému, registrace a služby RAGs, centrální databáze ISNI (počáteční import a sloučení záznamů z VIAF a 15 dalších specializovaných autorských DB – kontrakt s OCLC)
- **RAGs – registrační agentury:** aktuálně 14 (Bowker – fyzické osoby a vydavatelé; Ringgold – organizace, Bibliotheque Nationale de France) – služby koncovým uživatelům (přidělování ID), interface na RA
- **ISNI metadata set** – jméno, národnost, datum narození/úmrť, pohlaví, město, země, vztah k dalším ISNI-ID, atd.

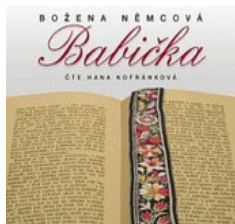
<http://www.isni.org/>

2.5 Trendy, další vývoj

- návrhy „chybějících“ id
image-id, licence-id, ...
- de-jure standardizace slibných de-facto id
MWLI (Musical Works Licence Id) aj.
- povinná referenční popisná metadata ke všem id
[identification + description](#)
- interoperabilita identifikátorů, **propojování**
book ISBN <-> ISTC, song ISWC <-> score ISMN <-> recordings ISRC
- **ISO TC46/SC9**
Technical Committee 46 – Standards of „Information and Documentation“
SubCommittee 9 – Presentation, identification and description of documents

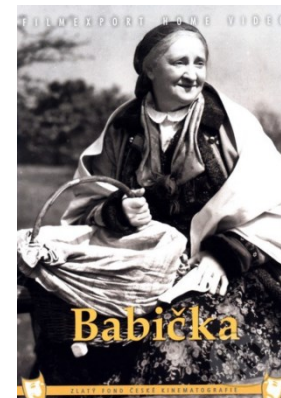
<https://www.iso.org/committee/48836.html>

Trendy – propojování

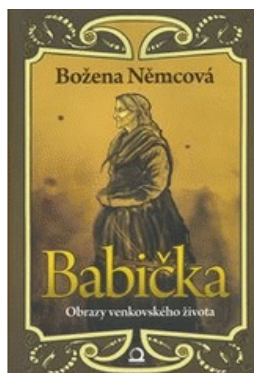


EAN: 8590236074224
CD/MP3 Radioservis 2014

ISTC: A12-2014-000135CD-E
Babička / Božena Němcová

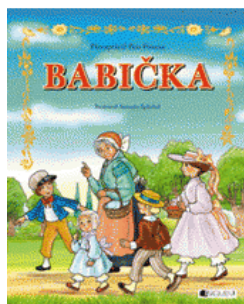


DVD, Filmexport
HomeVideo 2006

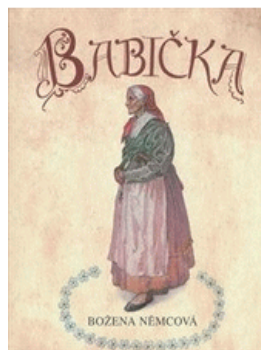


ISBN 978-80-7390-049-6
Praha, Dobrovský 2013

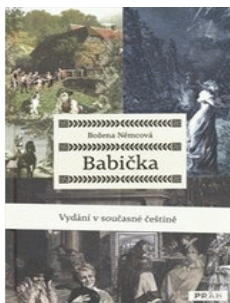
Filmový scénář 1940
Karel Hašler, Lucernafilm



ISBN 978-80-7459-021-4
Čs-spisovatel 2011



ISBN 978-80-242-2872-3
Knižní klub 2010



ISBN 978-80-7252-462-4
Praha, Práh 2013

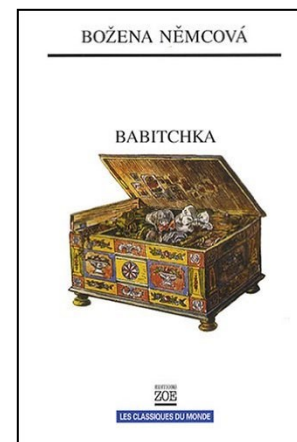


FI MU 2020

M. Bartošek - Digitální knihovny



ISBN 978-9989-163-94-4
Skopje 2009



ISBN 978-2-88182-628-3
Genève 2008

Digitální identifikátory



3. URN



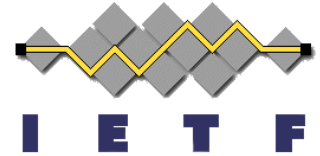
3. Inet a globální identifikace

Důležitost vhodného identifikačního systému pro zdroje v globální síti

- *eliminace fyzického kontaktu znamená vyšší potřebu identifikace (e-commerce)*
- *„document formats and protocols may come and go, but URIs will remain as the glue that binds the Web together“ (W3C, URI Activity Statement)*
- WWW : identifikátor **URL** (Uniform Resource Locator) – geniální, **ale** :
 - identifikuje zdroj prostřednictvím přístupového mechanismu (lokace)
 - při přesunu zdroje či změně systému -> mění se i identifikátor (1/6 URL každých 6 měs)
- potřeba trvanlivého (neměnného) identifikátoru **URN** již od počátku Internetu, **ale** :
 - dlouhá a trnitá cesta k němu - rozdílné přístupy (URN x lépe spravovaná URL)
 - kompetenční nejasnosti (IETF versus W3C)
 - dodnes není *prakticky* zcela dořešeno (syntaxe a funkčnost jasná, chybí implementace)
 - kdo zaplatí náklady na vytvoření a provoz potřebné infrastruktury ?

A **resource** can be anything that has identity. Familiar examples include an electronic document, an image, a service (e.g., "today's weather report for Los Angeles"), and a collection of other resources. Not all resources are network "retrievable"; e.g., human beings, corporations, and bound books in a library can also be considered resources. (RFC 2396)

3.1 URN



- **URN:nid:nss - URN:issn:0167-8191**
 - nid – *namespace*: id nějakého registrovaného identifikačního systému
 - nss - identifikátor objektu v rámci daného namespace
- URL – Uniform Resource **L**ocator
URN – Uniform Resource **N**ame } URI
- Analogie z knihoven: signatura (URL) x identifikátor ISBN (URN)
- URN = *zastřešení* pro různá identifikační schémata
- idea jasná od poč. 90 let (RFC 1737), **praxe zatím ne**
 - příliš náročná globální implementace

<http://www.w3.org/Addressing/>

4. PURL

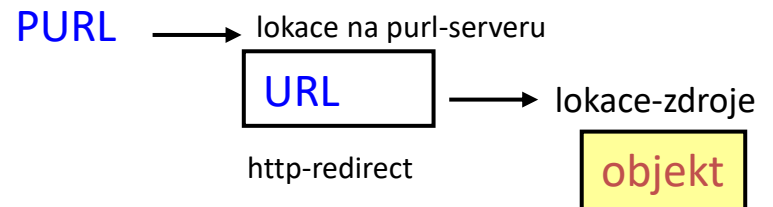


4. PURL – persistentní URL

- *Motto: využít URL, ale odstranit závislost na lokaci*
- 1. PURL je URL
- 2. PURL ukazuje na místo, kde je uloženo URL ukazující na objekt

- <http://purl.oclc.org/catalog/item1>

(na lokální adrese „/catalog/item1“
PURL serveru je uloženo URL zdroje)



- Pragmatické řešení firmy OCLC
(možnost instalovat vlastní PURL-server, open-source)

- <http://www.oclc.org/research/themes/data-science/purl.html>



5. Handles



5. handles

- [hdl:cnri.dlib/magazine](http://hdl.cnri.dlib/magazine) , <http://hdl.handle.net/10338.dmlcz/141708>
- technologie vyvinutá pro DL (viz Kahn-Wilensky architektura)
(**CNRI** – Corporation for National Research Initiatives, USA, od 1994)
- použita v NCSTRL, NDLP, DSpace, DOI, ...
- charakteristiky systému:
 - nezávislý na konceptu URN a systému DNS, ale „kompatibilní“
 - prověřený a fungující v řadě velkých praktických systémů
 - dobře škálovatelný
 - funkční mechanismus resoluce poskytující
 - rychlou resoluci
 - násobnou resoluci
 - otevřený standard a dostupná veškerá SW podpora
 - závazek dlouhodobého vývoje a udržování provozuschopnosti (**CNRI**)

5.1 handle ID

- handle system :
 - Identifikátory (nezávislé na lokaci)
 - resoluční systém (distrib. systém serverů s decentralizovanou administrací)
- **< naming authority >/< name >** **NASA.LaRC/tm112871**
 - *pojmenovací autorita* (NA) – oprávněna vytvářet a udržovat své handles *jméno* – jednoznačný id v rámci dané NA (lib. syntax)
 - hierarchický systém NA: **<NA>.<sub-NA>.<sub-sub-NA> ...**
 - globální NA (1.úrovně) přiděluje CNRI, sub-NA přiděluje nadřazená NA a všechny NA jsou registrovány v globálním serveru (registru) CNRI
- použití handles na straně klienta (www-prohlížeč) :
 - **www plug-in** pro přímou resoluci (přes Handle-protokol)

 - **resoluce přes proxy server** (již ne plná funkčnost)

<http://www.handle.net/>

5.2 Architektura hdl-systému

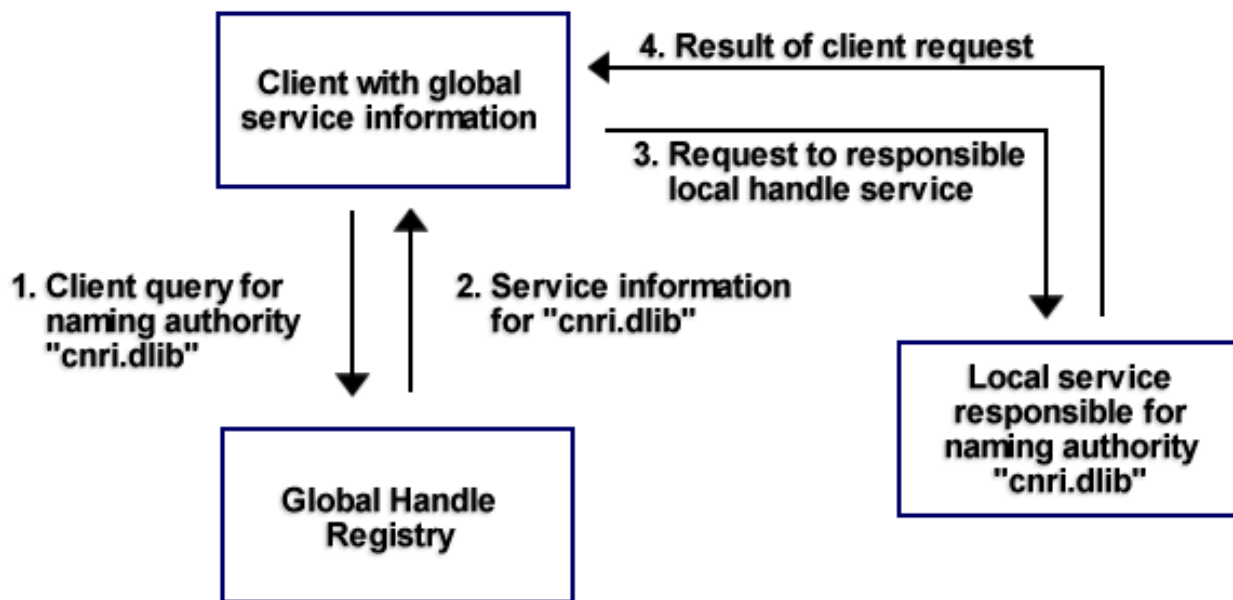
Globální server (Registr)

- pouze jeden v celém systému (CNRI)
- správa a resoluce jmen všech NA

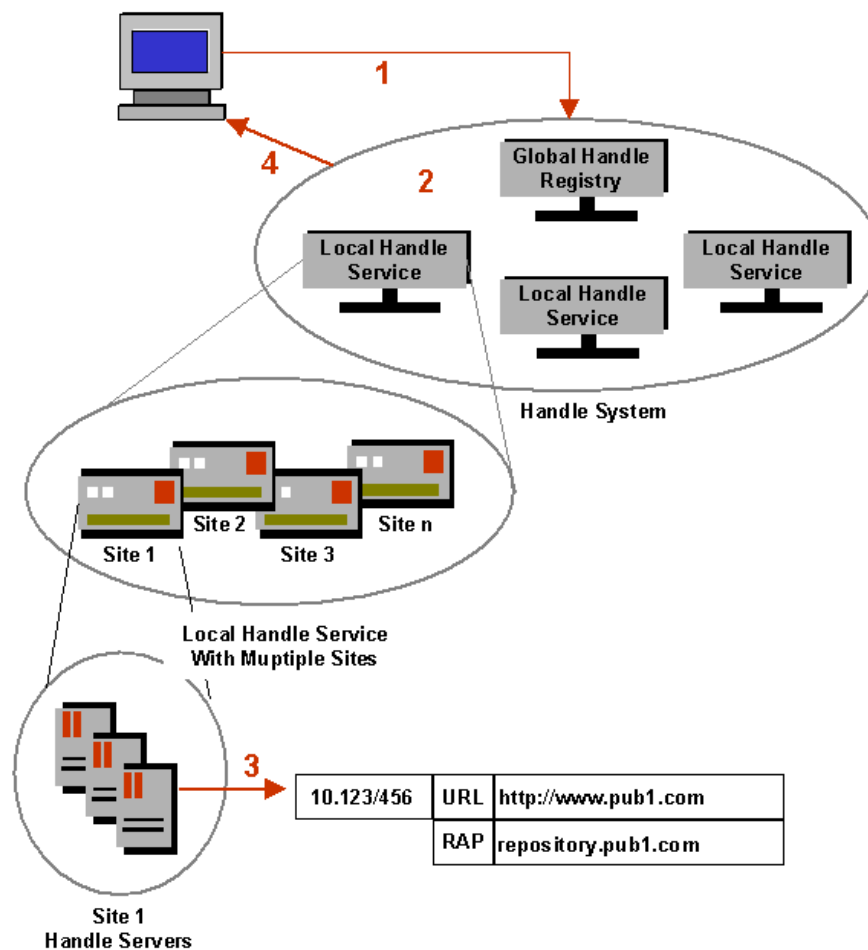
Lokální handle-servery

- resoluce a administrátorské služby pro lokální jména

- resoluce jména **hdl:cnri.dlib/july95-arms**



5.2 Architektura hdl-systemu



6. DOI



6. DOI – Digital Object Identifier

- **DOI:10.1006/123456**
- systém pro identifikaci jakýchkoliv entit – fyzických, digitálních, abstraktních (původně – copyrighitem chráněná díla, Asociace amerických nakladatelů, 1996)
- samofinancující se systém – otevřený, ale ne bezplatný (poplatek za přidělení DOI)
- velmi dynamicky se rozvíjející, s funkční resolucí
- Postaven na technologii **handle-system**
- **IDF – International DOI Foundation, 1998**
 - DOI pro identifikaci čehokoliv: dílo, projev, vyjádření, ..., publikace, data, ...
 - funkční od 2000, ISO standard od 2012
 - 2019/03: >190 miliónů přidělených DOI, >20.000 zapojených institucí
 - slibné, ale je to univerzální řešení ? (open ale ne free, handle)
 - nové systémy nad DOI – příklad **EIDR** (Entertainment ID – filmová a televizní díla)
- <http://www.doi.org/> skvělý Handbook: [doi:10.1000/182](https://doi.org/10.1000/182)
<http://dx.doi.org/10.1000/182>

6.1 Komponenty DOI-systému

DOI nejsou jen identifikátory, ale celý systém

enumeration	přiřazení čísel DOI (syntaxe, metodika jak-čemu přidělovat DOI, legacy id, ...)
description	vytvoření popisu/metadat (<i>indecs</i> framework)
resolution	číslo DOI → informace o DO (handle system)
policy	pravidla pro fungování systému (uživatel musí vždy dostat spolehlivý & předvídatelný výsledek)

Cílem: funkční, spolehlivý, persistentní samofinancující systém

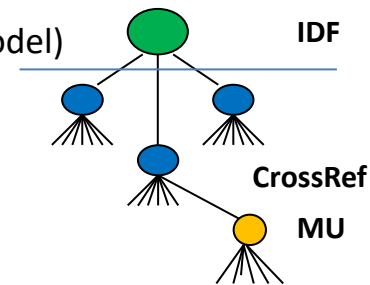
(vývoj a provoz infrastruktury musí vždy někdo nějak zaplatit;

persistence je vždy záležitostí služeb, nikoliv technologie !)

Platí se za přidělování čísel DOI (alá DNS domény), nikoliv za jejich používání

6.2 DOI policy (organizace)

- **RA – registrační autorita:** IDF – rozvoj systému a stanovení pravidel, registrace a služby pro RAGs, centrální databáze DOI, garance, financování infrastruktury
- **RAGs – registrační agentury** (aktuálně 10, např. CrossRef, DataCite, ...)
 - přidělování DOI-prefixů registrujícím organizacím (**10.5817 = MU** u CrossRef)
 - služby registrujícím organizacím a koncovým uživatelům, specializace v určité oblasti (CrossRef = vědecké články, DataCite = vědecká data, EIDR = filmy a TV pořady, ...)
 - dva „nezávislé“ obchodní modely
 - poplatky RAG -> RA : franchise fee (fixní roční poplatek)
 - poplatky RO -> RAG : registration fee (každá RAG vlastní nezávislý obchodní model)
- **RO – registrující organizace** (přes 16.000)
 - vydavatelé, univerzity, knihovny aj. s přiděleným DOI prefixem
 - registruje čísla DOI a metadata pro „své“ objekty u své mateřské RAG
 - zodpovídá za aktualizaci stavová data pro svá DOI (aktuální URL)
 - MU – zaregistrována u CrossRef: příklad: DOI **10.5817/AM2013-1-17**
resoluce: <https://doi.org/10.5817/AM2013-1-17>



6.3 DOI enumeration

prefix

sufix

doi:10.1000/ISBN-1-900512-44-0

doi:10.5817/AM2013-1-17

- **prefix:** jednoznačný číselný identifikátor registrující organizace
(10. = označení aplikace DOI v rámci systému handles)
- **sufix:** jednoznačná identifikace entity v rámci RO
(libovolný lokální identifikační systém, lokálně jednoznačný)
- DOI číslo nepředpokládá žádnou sémantiku (dumb)
- velikost či struktura prefixu i sufixu neomezena
- registrující organizace může mít lib. počet prefixů
- prefixy číselné (nepovinné, názvy vydavatelů se často mění)
- číslo DOI se nemění při změně práv (vlastníka objektu)

6.4 DOI resolution

- resoluce prostřednictvím *handle-systém* (DOI = aplikace nad handle)
- dvě metody pro odkazování čísel DOI na Internetu:
 - DOI jako URI: doi:10.123/456
 - DOI jako URL: <https://doi.org/10.123/456> (<https://doi.org/{DOI}>)
- dále platí totéž, co pro handle – resoluce DOI přes www-prohlížeč:
 - přímo (plug-in schopný zpracovávat handle-protokol)
 - přes URL a proxy (gateway mezi HTTP a handle-protokolem)

References:

- [1] Balinsky, A., Ryan, J.: *Some sharp L^2 inequalities for Dirac type operators*. SIGMA, Symmetry Integrability Geom. Methods Appl. (2007), 10, paper 114. [DOI 10.1037/9871-3520.5.98](https://doi.org/10.1037/9871-3520.5.98)
- [2] Brackx, F., Delanghe, R.: *Clifford Analysis*. Pitman, London, 1982. [DOI 10.2100/BDC398762](https://doi.org/10.2100/BDC398762)
- [3] Bureš, J., Sommen, F., Souček, V., Van Lancker, P.: *Rarita-Schwinger type operators in Clifford analysis*. J. Funct. Anal. 185 (2) (2001), 425–455. [DOI 10.1105/i-funct-anal.2001.425](https://doi.org/10.1105/i-funct-anal.2001.425)

6.5 DOI – CrossRef



- **CrossRef** = registrační agentura **pro oblast vědeckých publikací**
 - jednoznačná identifikace a propojování vědeckých článků (**citation linking**)
 - 12.000 RO, přes 108 mil DOI , <http://crossref.org>
 - DOI resolver: <http://dx.doi.org/>
- **Povinnosti RO** (např. MU)
 - Webová verze publikace (např. časopisu)
 - response page: www-stránka, na kterou je směřováno přidělené DOI
 - **Inbound linking**: přidělování DOI vlastním vydaným článkům
 - registrace DOI a požadovaných metadat v databázi CrossRef
 - průběžná aktualizace metadat
 - **Outbound linking**: dohledání a dopsání DOI k článkům odkazovaným v seznamu literatury u daného článku
 - DOI jako součást každé reference = propojená síť odkazů na webu
 - poplatky CrossRef: roční fixní + 1 USD/1 DOI (0,15 USD za archivní DOI)

6.5 DOI – CrossRef



Czech Digital Mathematics Library

[About DML-CZ](#) | [FAQ](#) | [News](#) | [Conditions of Use](#) | [Math Archives](#) | [Contact Us](#)

[Previous](#) | [Up](#) | [Next](#)

Search

[Advanced Search](#)

Browse

- [Collections](#)
- [Titles](#)
- [Authors](#)
- [MSC](#)

[About DML-CZ](#)

Partner of



[DML-CZ Home](#) >
[Archivum Mathematicum](#) >
[Volume 48](#) >
[Issue 3](#) >

Article

[LAOHAKOSOL, VICHIAN](#) ; [PIMSERT, WATCHARAPON](#) ; [HENGKRAWIT, CHARINTHIP](#) ; [EBANKS, BRUCE](#)

Some logarithmic functional equations. (English). Archivum Mathematicum, vol. 48 (2012), issue 3, pp. 173-181
MSC: [39B20](#) | [MR 2995870](#) | DOI: [10.5817/AM2012-3-173](#)

[Full entry](#) | [PDF](#) (0.4 MB)

[Feedback](#)

Keywords:

logarithmic functional equation; Pexider equations

Summary:

The functional equation $f(y-x) - g(xy) = h\left(\frac{1}{x-1/y}\right)$ is solved for general solution. The result is then applied to show that the three functional equations $f(xy)=f(x)+f(y)$, $f(y-x)-f(xy)=f\left(\frac{1}{x-1/y}\right)$ and $f(y-x)-f(x)-f(y)=f\left(\frac{1}{x-1/y}\right)$ are equivalent. Finally, twice differentiable solution functions of the functional equation $f(y-x) - g_1(x)-g_2(y) = h\left(\frac{1}{x-1/y}\right)$ are determined.

[Similar articles:](#) [?](#)

References:

- [1] Chung, J.-Y.: **A remark on a logarithmic functional equation.** J. Math. Anal. Appl. 336 (2007), 745–748. [DOI 10.1016/j.jmaa.2007.02.072](#) | [MR 2348539](#) | [Zbl 1130.39018](#)
- [2] Ebanks, B.: **On Heuvers' logarithmic functional equation.** Result. Math. 42 (2002), 37–41. [DOI 10.1007/BF03323552](#) | [MR 1934223](#) | [Zbl 1044.39018](#)
- [3] Heuvers, K. J.: **Another logarithmic functional equation.** Aequationes Math. 58 (1999), 260–264. [DOI 10.1007/s000100050112](#) | [MR 1715396](#)
- [4] Heuvers, K. J., Kannappan, P.: **A third logarithmic functional equation and Pexider generalizations.** Aequationes Math. 70 (2005), 117–121. [DOI 10.1007/s00010-005-2792-8](#) | [MR 2167989](#) | [Zbl 1079.39019](#)

7. ARK



7. ARK – využití URL

ARK = Archival Resource Key (California Digital Library)

- **Link Rot:** problém není v URL ale v tom, jak moc vážně bere poskytovatel inf-zdrojů svůj závazek udržet je dlouhodobě dostupné (nejedná se o problém technologický, ale administrativní, organizačně-sociální)
- persistence *je vždy složitá a drahá záležitost*, kterou vhodná technologie může usnadnit, ale sama o sobě ji neřeší
- jestliže se organizace nestará o přesměrovávání svých URL (což je relativně snadný a dobře zvládnutelná proces) dnes, těžko se z ničeho nic začne starat o údržbu stavových dat svých identifikátorů URN či Handles (což je navíc záležitost poměrně nová a ne moc prozkoumaná)
- přístupy založené na využití nepřímé adresace nic neřeší; pracnost udržování aktuálnosti tabulek pro nepřímou adresaci je stejná jako udržování URL a jejich přesměrovávání
- **závěr: nepotřebujeme URN ale rozumnou koncepci využívání URL a disciplinu při jejím dodržování !**

7.1 ARK = udržované URL

<http://ark.cdlib.org/ark:/13030/ft4w10060w>

<http://NMAH/ark:/NAAN/Name>

- oddělení 2 různých služeb (přidělování id , garantování dostupnosti objektu)
- **NAAN** – Name Assigning Authority Number
 - NAA zajišťuje počáteční přidělení identifikátoru
 - má globálně jednoznačný číselný kód, který je součástí ARK-id
 - NAA může časem zaniknout, NMA existuje stále (může se změnit)
- **NMAH** – Name Mapping Authority Host
 - NMA zajišťuje trvalou dostupnost objektu – směřuje identifikátor
 - neslouží k identifikaci či při porovnávání ekvivalence dvou ARK-id
 - může se časem změnit (resp. jich může být i více pro jeden ARK-id)
 - globální DB: která NMAH obsluhuje ARKs vydané danou NAA?
- <http://ark.cdlib.org/ark:/13030/ft4w10060w> je ekvivalentní s
<http://some.other.org/ark:/13030/ft4w10060w> je ekvivalentní s
<ark:/13030/ft4w10060w>

7.2 ARK funkcionalita

- URL = vazba na jeden objekt
- ARK = vazba na **tři** různé věci:
 - **ARK** – vrací objekt
 - **ARK?** – vrací metadata objektu
 - **ARK??** – vrací „commitment statement“
- http://n2t.net/e/ark_ids.html
- <https://tools.ietf.org/id/draft-kunze-ark-21.html>

7.3 Formát identifikátorů

- **Žádnou sémantiku do identifikátorů !!**
 - jazykově založená sémantika
 - akronymy, zkratky
 - jakákoliv další sémantika v identifikátoru
 - věci se časem mění, získávají jiný význam
 - = hrozba pro persistenci
- snaha vyhnout se i *neúmyslné* sémantice
 - pouze číslice a souhlásky
 - ne více jak dvě souhlásky vedle sebe
 - kontrolní znak
 - náhodné generátory pro předcházení **sériové sémantice** (možnost odhadnout pořadí přidělení identifikátorů)

Literatura



Doplňková literatura

- H.W.Hilse, J.Kothe. *Implementing Persistent Identifiers*. (CERL 2006)
http://webdoc.sub.gwdg.de/edoc/ah/2006/hilse_kothe/urn%3Anbn%3Ade%3Agbv%3A7-isbn-90-6984-508-3-8.pdf
- G.Vitiello. *Identifiers and Identification Systems*. (Dlib Jan 2004)
<http://www.dlib.org/dlib/january04/vitiello/01vitiello.html>
- J.Hakala. *Persistent identifiers – an overview*. TWR Technology Watch Review, 2010. <http://www.persid.org/downloads/PI-intro-2010-09-22.pdf>
- ISO standardizační proces:
http://www.iso.org/iso/home/standards_development.htm