

Elektronické informační zdroje

Technologie a protokoly v EIZ

25. 10. 2019

Co už víme...

- co jsou to EIZ
- jak se nakupují komerční EIZ
- jaké jsou ty významné
- jak se vyhledává (v základech)
- jak se to všechno propojí

Mgr. Tomáš Marek
marek@kisk.cz

Co nás dnes čeká?

- co jsou to EIZ
- jak se nakupují komerční EIZ
- jaké jsou ty významné
- jak se vyhledává (v základech)
- jak se to všechno propojí
- co je ještě za tím vším?

Mgr. Tomáš Marek
marek@kisk.cz

<http://sk.sagepub.com/cqpress/100-americans-making-constitutional-history?fromsearch=true>

Identifikátory

- Jaké tradiční identifikátory znáte?
- Co všechno potřebujeme jednoznačně identifikovat?

Tradiční identifikátory

ISBN – International Standard Book Number

ISSN – International Standard Serial Number

ISMN – International Standard Music N. for Printed Music

ISAN – International Standard Audiovisual Number

ISRN – International Standard Technical Report Number

Ani v digitálním prostředí neztrácejí smysl...

ISBN



- International Standard Book Number
- vznik 1966, agentura ISBN v Berlíně
- u nás řeší/vydává NK v Praze
- ISBN-10 a ISBN-13 (od 2007) – EAN13
- UCC 978 *Bookland* (+979, dříve *Musicland*)
- ISBN v elektronickém prostředí
- www.isbn-international.org

ISBN syntaxe



ISBN (978)-80-00-01987-6

- identifikátor skupiny
- identifikátor nakladatele
- identifikátor titulu
- kontrolní číslice

- <https://www.isbn-international.org>
- Global Register of Publishers (978-80-87989-06-7)

ISSN



INTERNATIONAL
STANDARD
SERIAL
NUMBER
CZECH REPUBLIC

- seriálové publikace
- 60. léta, agentura ISSN v Paříži
- u nás řeší/vydává NTK v Praze
- osmičíselný „hloupý“ identifikátor
- navázán na jmenné údaje (změna názvu)
- www.issn.cz / www.issn.org
- http://aleph.ntkcz.cz/F/?func=find-b-0&local_base=stk02

ISSN

- p-ISSN, e-ISSN, ISSN-L
- <https://portal.issn.org/>
- Vesmír
- <https://vesmir.cz/cz/>



INTERNATIONAL
STANDARD
SERIAL
NUMBER
CZECH REPUBLIC

DOI

- Digital Object Identifier
- jednoznačná dohledatelnost dokumentu (**objektu**) na webu
- iniciativa komerčních vydavatelů
- snaha o vybudování komplexního systému na správu a řízení vlastnických a autorských práv
- centralizovaný, placený systém
- <https://www.crossref.org/fees/>



DOI Syntaxe



doi:10.1006/123456

- konstanta pro označení systému DOI
- numerický ID kód registrující organizace, vydavatele
- jednoznačný identifikátor DO v rámci dané reg. organizace

- prefix přiděluje registrační agentura
- sufix je libovolný na základě pravidel vydavatele
- DOI se uděluje pouze vydaným dokumentům

DOI Syntaxe



- 10.1000/182
Manuál mezinárodní nadace DOI
- 10.1145/336597.336660
stať z konf. sborníku publikovaná spol. ACM
- 10.1016/j.acalib.2006.04.004
článek z časopisu Journal of Academic Librarianship
- DOI Resolver: <http://www.doi.org/>

DOI varianty

- 10.978.12345/99990



DOI varianty

- 10.978.12345/99990
- 10.5930/issn.1994-4683



DOI varianty



- 10.978.12345/99990
- ISBN-A
- *actionable* ISBN
- *„using DOI multiple resolution features, the ISBN may be turned into a one-click gateway to enhance the ,book experience“*
- přesměrování např. na web o dané knize
- print-on-demand, volba formátu atp.
- <http://www.doi.org/factsheets/ISBN-A.html>

CrossRef



- nejvýznamnější registrační agentura
- nadstavbové nástroje:
CrossCheck, Cited-by Linking
- LAB: <http://crossref.github.io/doi-popup/>
- <http://chronograph.labs.crossref.org/>
- <http://shortdoi.org/>

Ještě nám něco chybí...

<https://academic.oup.com/ahr/article/124/4/1267/5581082>

ORCID

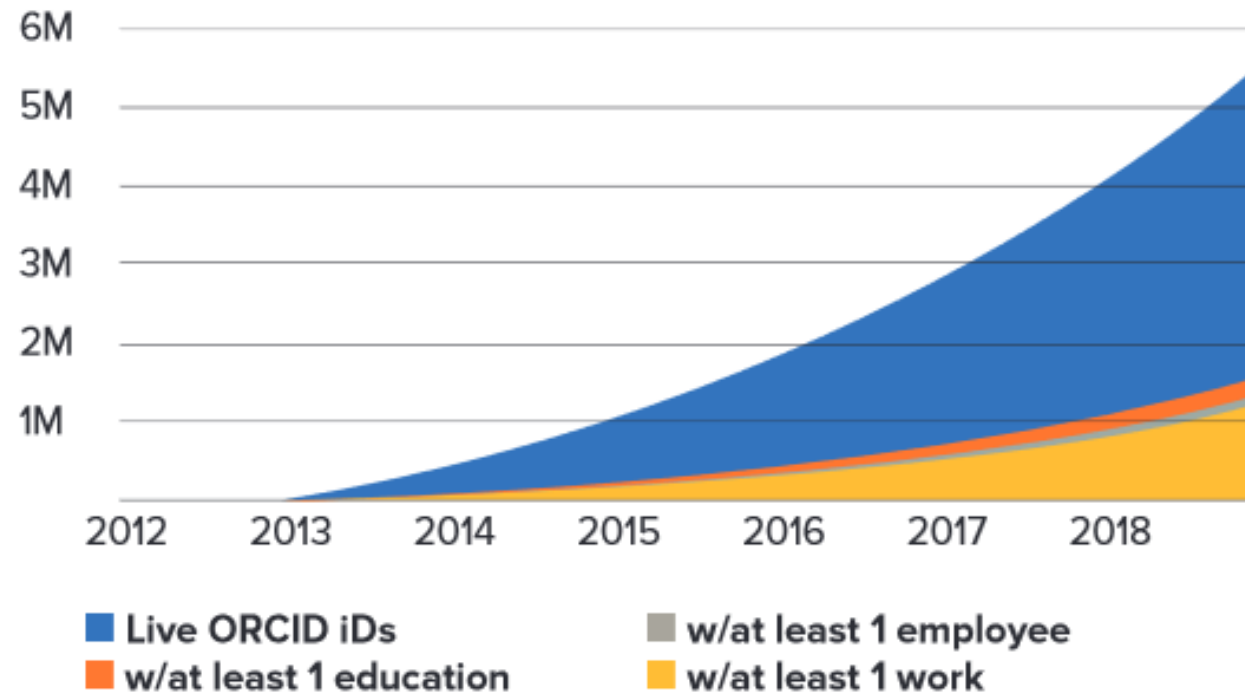
- potíže se jmény (neunikátní, měnící se, různá)
- identifikátor autora/výzkumníka
- <https://orcid.org/>
- 0000-0002-1825-0097
- <https://journals.sagepub.com/toc/ivia/current>



ORCID
Connecting Research
and Researchers

ORCID

Connections to ORCID records



Source: [Orcid.org/statistics](https://orcid.org/statistics) as of November 2018

- **3.7** ORCID iDs created every minute
- **193** different languages included in ORCID records
- **73.6%** of records have granted update permissions
- **1.97** record updates made per second

Metadatové formáty

- Značkovací jazyky obecně
- HTML
- *Markdown*
- XML
- MarcXML
- DC
- JSON

XML



- eXtensible Markup Language
- obecný značkovací jazyk
- vyvinut W3C jako nástupce SGML
- základem pro konkrétní značkovací jazyky
- definuje strukturu dokumentu z hlediska věcného
- nezabývá se vzhledem
- vzhled se definuje připojeným stylem

Využití XML



- prezentace informací
- ukládání informací
- výměna informací
- popis informací (metadata)
- získávání informací

XML v praxi



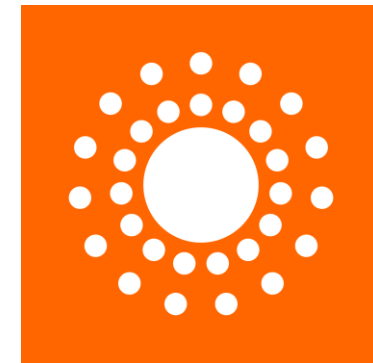
- specifikace volně: <https://www.w3.org/TR/REC-xml/>
- SVG – *Wiki SVG*
- RSS
- KML
- MS Office a Open Office
- MarcXML
- ...

MarcXML

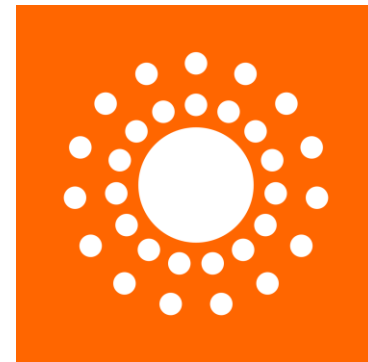
- standard pro uchovávání záznamů MARC21 v XML
- původně pro potřeby knihoven
- správce LoC
- <http://www.loc.gov/standards/marcxml>
- <https://krizovatka.koha.cloud/cgi-bin/koha/opac-detail.pl?biblionumber=20059>

Dublin Core / DCMI




- soubor metadatových prvků
- pro popis digitálních objektů (i webů)
- usnadňuje vyhledávání e-zdrojů
- založen na XML
- [Generátor metadat](#)
- název odvozen od města Dublin (USA)
- <http://www.dublincore.org>
- http://www.ics.muni.cz/dublin_core



Dublin Core / DCMI



- Dublin Core Metadata Element Set

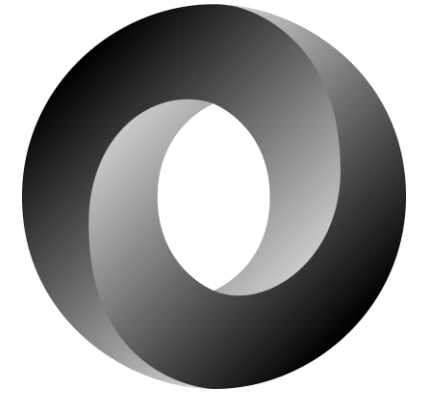
	dc.creator	Ediciones El País	
	dc.date	2014-07-11	
	dc.description	Noticias de última hora sobre la actualidad en España y el mundo: política, economía, deportes, cultura, sociedad, tecnología, gente, opinión, viajes, moda, televisión, los blogs y las firmas de EL PAÍS. Además especiales, vídeos, fotos, audios, gráficos, entrevistas, promociones y todos los servicios de EL PAÍS.	
	dc.language	es	
	dc.publisher	Ediciones El País	
	dc.subject	noticias, última hora, actualidad, nacional, España, ccaa, internacional, política, economía, deportes, cultura, sociedad, tecnología, gente, opinión, viajes, moda, televisión, blogs, firmas, especiales, vídeos, fotos, audios, gráficos, entrevistas, servicios	
	dc.title	EL PAÍS: el periódico global	

- <https://schema.org/>

Další metadatové formáty

- METS (Metadata Encoding and Transmission Standard)
- MODS (Metadata Object Description Schema)
- kompromis mezi jednoduchým DC a složitým MARC
- <http://www.digitalniknihovna.cz>
- a mnoho dalších...

JSON



- JavaScript Object Notation
- odlehčený formát pro výměnu dat na webu
- jednoduše čitelný i zapisovatelný člověkem
- snadno analyzovatelný i generovatelný strojově
- v posledních letech jeden z nejvýznamnějších
- <https://catalog.data.gov/dataset>
- *JSON plugins*

users.xml

```
<users>
  <user>
    <username>SammyShark</username> <location>Indian Ocean</location>
  </user>
  <user>
    <username>JesseOctopus</username> <location>Pacific Ocean</location>
  </user>
  <user>
    <username>DrewSquir</username> <location>Atlantic Ocean</location>
  </user>
  <user>
    <username>JamieMantisShrimp</username> <location>Pacific Ocean</location>
  </user>
</users>
```

users.json

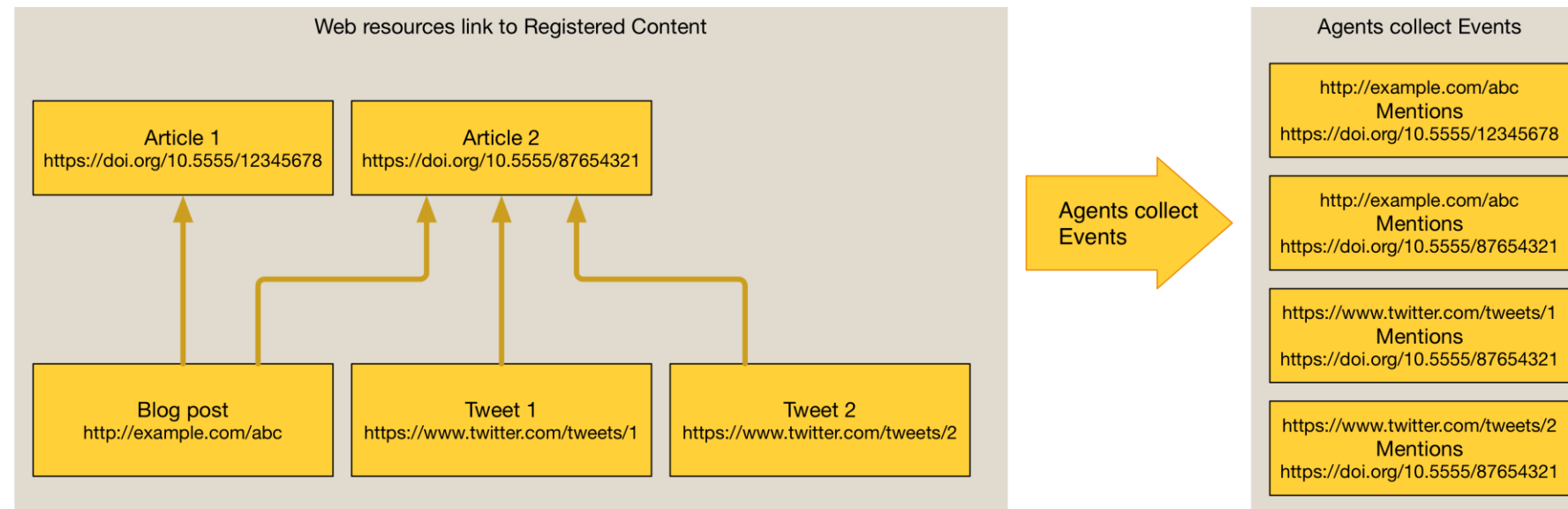
```
{ "users": [
  { "username" : "SammyShark", "location" : "Indian Ocean"},
  { "username" : "JesseOctopus", "location" : "Pacific Ocean"},
  { "username" : "DrewSquid", "location" : "Atlantic Ocean"},
  { "username" : "JamieMantisShrimp", "location" : "Pacific Ocean"}
] }
```

CrossRef



Event Data API

<https://api.eventdata.crossref.org/v1/events?obj-id=>



https://en.wikipedia.org/wiki/Digital_object_identifier

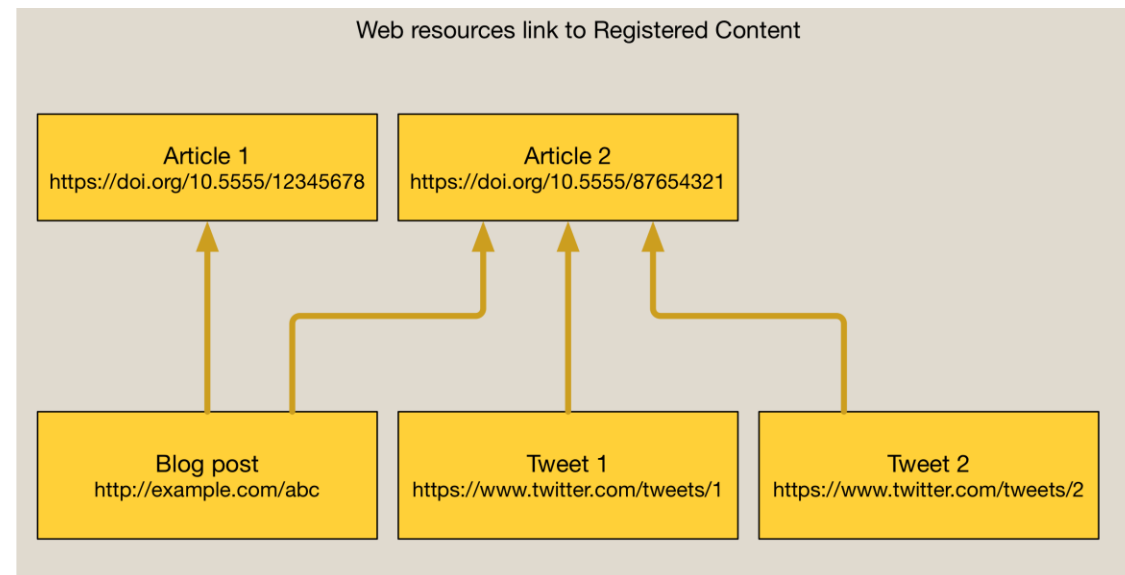
<https://www.eventdata.crossref.org/guide/service/query-api/>

CrossRef



Event Data API

<https://api.eventdata.crossref.org/v1/events?obj-id=>



Agents collect Events



https://en.wikipedia.org/wiki/Digital_object_identifier
<https://www.eventdata.crossref.org/guide/service/query-api/>

RSS

- *RDF Site Syndication / Rich Site Summary*
- standardizovaný formát (XML)
- pro weby s častou aktualizací
- zpravodajské servery, blogy
- [osobní RSS čtečky](#)
- agregátory RSS kanálů
- RSS vs Atom



K čemu je to dobré?

- informace o novinkách
- aktuální čísla periodik
- aletry na vybrané klíčové slovo nebo obor

Current Awareness Services

- <http://www.tictocs.ac.uk>



Protokoly

- Protokoly k prohledávání vzdálených IZ
- *Z39.50, SRU/SRW*

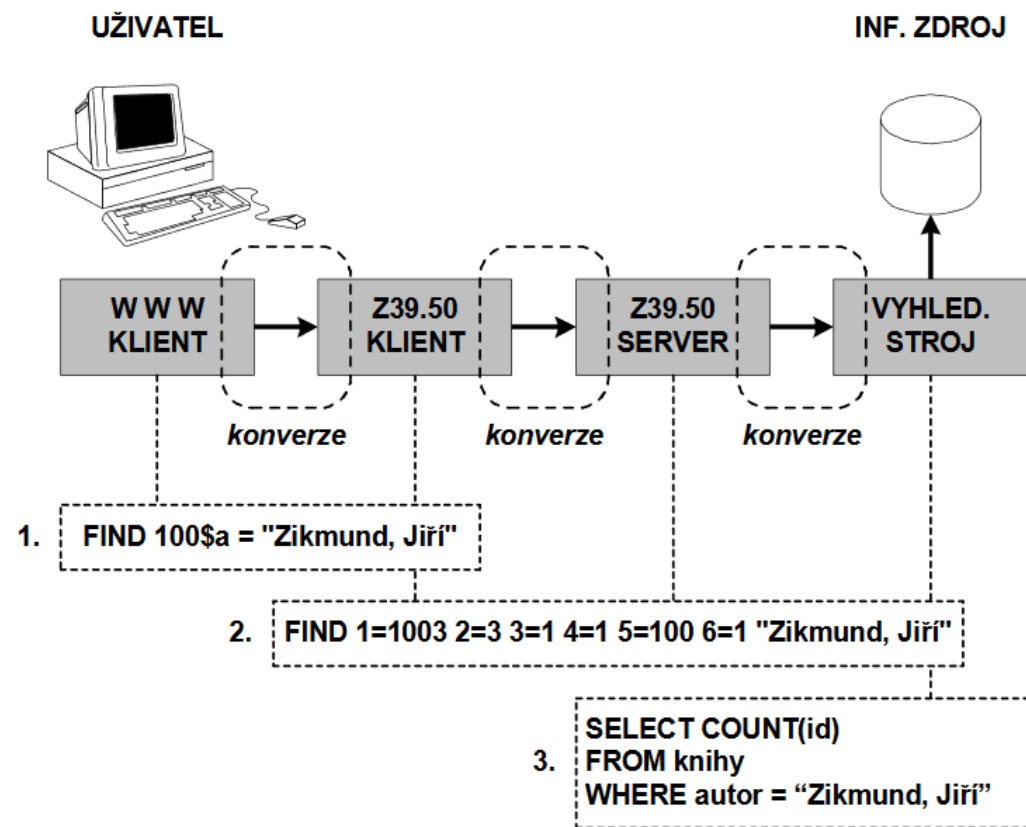
- Protokoly ke sklizení vzdálených IZ
- *OAI-PMH*

Z39.50

- mezinárodním standard pro komunikaci mezi počítači
- architektura klient – server
- heterogenní prostředí - nezávislé na operačních systémech, databázích a dotazovacích jazycích
- rozvoj protokolu řídí mezinárodní skupina Z39.50 *Implementors Group* (ZIG), patronem LoC
- **před-Webová technologie**, snahy o modernizaci

Fungování Z39.50

- uživatel - dotaz v jazyce svého systému
- cizí vzdálený katalog se Z-serverem
- dotaz přeformulován do Z39.50
- pošle se Z-serveru cizího katalogu
- dotaz se přeloží do jazyka cílového IZ
- vrátí se výsledek vyhledávání
- výsledek pošle Z-klientovi
- Z-klient výsledek předá systému
- zobrazení v uživatelském rozhraní



SRU/SRW

- protokol pro online vyhledávání
- pokus o zjednodušení Z39.50
- využití osvědčených webových technologií (HTTP, XML)

- SRU - *Search/Retrieve via URL*
- jednodušší implementace než Z39.50
- vyvíjí a spravuje LoC, myšlenka vychází z openURL

SRU

- dotaz vyjadřuje jazykem CQL a kóduje ho do URL
- URL obsahuje adresu SRU serveru + CQL dotaz v předepsané syntaxi
- adresa **?** vyhledávací část **&** parametr (klíč=hodnota)
- <http://z3950.loc.gov:7090/voyager?version=1.1&operation=searchRetrieve&query=brno>
- <https://www.loc.gov/standards/sru/misc/simple.html>

SRW

- Search/Retrieve Web Service
- obdoba SRU
- komunikace z obou stran skrze XML

K čemu je to dobré?

Z39.50

SRU/SRW

- komunikace mezi systémy
- prohledávání vzdálených IZ
- budování virtuálních katalogů
- federalizované centrální vyhledávání

OAI-PMH

- protokol pro automatické sklízení dat
- centrální ukládání dat
- Open Archives Initiative Protocol for Metadata Harvesting
- <https://www.openarchives.org/pmh/>



OAI-PMH



- tvorba citačních rejstříků
 - [centrální indexy](#) virtuálních repozitářů
 - [Třeba taky k tomuhle...](#)
-
- ResourceSync Framework Specification
 - Open Archives Initiative Object Reuse and Exchange (OAI-ORE)

Takže...

- Projděte si všechno klikací
- Procházejte EIZ a hledejte různé formáty
- Nastavte si RSS odběr v EIZ a podívejte se, jak to vypadá
- Zkuste si dotazovat API CrossRefu
- Stáhněte se z libovolného EIZ JSON
- Nainstalujte si doplněk na JSON formátování