

DOI/CrossRef Metadata Schema

Masarykova Univerzita
Fakulta Informatiky

Ondřej Ryšánek, 3. ročník,
obor Umělá inteligence
a zpracování přirozeného jazyka

27. listopadu 2019

URL: <https://www.crossref.org/>

1 DOI

Identifikátor digitálního objektu (*angl. Digital Object Identifier, zkr. DOI*) je unikátní a trvalý identifikátor děl přístupných v digitální verzi. Roku 1996 jej vytvořila Asociace amerických nakladatelů, ale od roku 1998 jej spravuje a rozvíjí nadace IDF (International DOI Foundation).

2 CrossRef

CrossRef je největší nezisková mezinárodní registrační agentura umožňující svým členům využívat DOI pro svoje odborné publikace. Byla založena ve spolupráci s IDF v roce 2000 skupinou vydavatelů, kteří potřebovali účinný a trvalý způsob, jak se ve svých online publikacích odkazovat na jiné publikace napříč různými vydavateli. Jejím členem se může stát jakákoliv instituce, která vydává vědeckou literaturu dostupnou v digitální formě. Každý takový vydavatel platí roční členský poplatek a dále také malou částku za každé přidělené DOI.

CrossRef nabízí další nadstavbové služby jako

- Cited-by — služba umožňující nalezení článků citujících momentálně zobrazenou publikaci
- Crossmark — získání změny stavu publikace ve smyslu její aktualizace či korekce a přístup k novým metadatům

- Event Data — shromažďování a zpřístupnění dat o komentářích, odkazech, sdíleních, zmínkách apod. týkajících se určité publikace
- Similarity Check — nástroj na kontrolu plagiátorství; zpoplatněn
- Funder Registry — databáze organizací nabízejících granty

Tyto služby pracují na základě dostupných metadat všech registrovaných publikací.

3 CrossRef Metadata

Při registraci své online publikace přiloží vydavatel metadata, jejichž součástí jsou mj. adresa umístění díla a přidělený DOI. Metadata se mohou případně doplňovat a měnit.

CrossRef metadata zahrnují

- Bibliografická metadata
- Metadata financování
- Licenční metadata
- Metadata klinických studií
- Vztahy mezi DOI navzájem a mezi DOI a jinými publikacemi nedigitální podoby
- Doplnění nebo odstranění obsahu publikací či metadat
- Abstrakty označované pomocí JATS¹

CrossRef metadata jsou popsána jazykem XML. Nejnovější verze XML schématu pro bibliografii nese název [crossref4.4.2.xsd](#). Počínaje touto verzí budou všechna další vydání dostupná také na webovém repozitáři [GitLab](#).

3.1 Bibliografická metadata

Nejsvrchnějším elementem při značkování metadat je *doi_batch*, který se dělí na hlavičku v elementu *head* a tělo v element *body*.

Hlavička se dále dělí na *doi_batch_id*, *timestamp*, *depositor* a *registrant*. Tyto informace jednoznačně identifikují příspěvek do databáze CrossRefu.

¹ Journal Article Tag Suite – schéma definující skupinu XML značek pro metadataový popis vědeckých článků; více na <https://jats.nlm.nih.gov/>

Tělo může obsahovat libovolný počet publikací. Jedinou podmínkou je, aby všechny byly stejného druhu. XML značky pro popis bibliografických údajů se mohou různit podle toho, o jaký typ obsahu se jedná.

CrossRef rozlišuje tyto druhy vědeckých děl

- Časopisecký článek
- Kniha / knižní kapitola
- Příspěvek v konferenčním sborníku
- Datová sada
- Peer review
- Zveřejněný nepublikovaný obsah
- Report / rozepsaná práce
- Standard
- Komponenta (obrázek, graf, tabulka, apod.)

3.1.1 Časopisecký článek

Metadata k časopiseckému článku se vkládají do elementu *journal*. Ten umožňuje vyjádřit metadata týkající se jak článku, tak časopisu, ve kterém byl publikován.

Časopisecká metadata zahrnují plný název a jeho zkratku, ISSN², CODEN³, lokace archivů a DOI. Také lze zaznamenat údaje ke konkrétnímu vydání časopisu, např.: přispěvatele, název i podnázev, datum publikace, ročník a číslo, lokace archivů a DOI.

V rámci časopisu lze samozřejmě také uvést bibliografické údaje k jednotlivým článkům, které obsahuje. Tyto údaje rovněž zahrnují název, podnázev, přispěvatele, datum publikace, lokace archivů a DOI. Dále jsou rozšířeny o abstrakt psaný pomocí JATS, datum přijetí k publikování, čísla stránek výskytu článku, identifikátory článku v rámci vydavatele, metadata služby Crossmark nebo trojici dat o licenci, financování či klinických studiích. Důležitou součástí údajů článku jsou i seznamy citovaných zdrojů a doplňujících materiálů.

² International Standard Serial Number – jednoznačný identifikátor tištěných i digitálních periodik; více na <https://www.issn.org/>

³ šestiznakový bibliografický identifikátor periodických i neperiodických publikací ze všech oborů; více na <https://en.wikipedia.org/wiki/CODEN>

3.2 Metadata financování

Metadatové schéma k popisu financování publikace se nachází v souboru [fundref.xsd](#), který už je importován do schématu [crossref4.4.2.xsd](#).

Kořenovým elementem při popisu financování publikovaného materiálu je element *program* s fixním atributem *name* nastaveným na hodnotu *fundref*. Do kořene lze vnořit libovolný počet elementů *assertion*, které mají atribut *provider* nabývající hodnot buď *publisher* nebo *crossref* a povinný atribut *name* s možnými hodnotami *fundgroup*, *funder_identifier*, *funder_name* nebo *award_number*.

Element *assertion* se může zanořovat sám do sebe, ale je třeba dávat pozor na pořadí dle atributu *name*. Hodnota *fundgroup* umožňuje seskupovat informace o financování publikace z více zdrojů. Hodnota *funder_identifier* jednoznačně identifikuje financující agenturu a musí být vnořena uvnitř elementu *assertion* s atributem *name="funder_name"*. Hodnota *funder_name* označuje jméno financující agentury. Poslední hodnota *award_number* slouží jako identifikátor grantu.

3.3 Licenční metadata

Stejně jako metadatové schéma pro financování je i schéma pro popis licencí nacházející se v souboru [AccessIndicators.xsd](#) importováno přímo do schématu [crossref4.4.2.xsd](#).

Kořenovým elementem při značkování licencí je element *program* s fixním atributem *name* nastaveným na hodnotu *AccessIndicators*. Do něj je vnořen nepovinný element *free_to_read* s nepovinnými atributy *end_date* a *start_date* a element *license_ref* obsahující URL kýžené licence. Pomocí volitelných atributů *start_date*, resp. *applies_to* elementu *licence_ref* se nastavuje datum začátku platnosti licence, resp. zda se licence váže k momentální verzi publikace, k verzi článku, který prošel skrz peer review a byl schválen k publikaci, anebo se licence vztahuje na *text and data mining*.

4 Vlastní hodnocení

Vyjadřovací sílu schématu hodnotím velmi kladně. Lze jím popsat dostatečně širokou škálu typů vědeckých prací. Nejvíce jsem se ve svém průzkumu zaměřil na vědecké články do časopisů, takže se nemůžu vyjádřit ke všem druhům publikací, ale myslím si, že veškeré důležité bibliografické i nebibliografické údaje se dají pomocí schématu nějak označkovat.

Množství značek se mi nezdá přemrštěné, autoři skutečně naimplementovali jen to nejpotřebnější. Díky této úspornosti se dá schéma vcelku rychle nastudovat, takže je jeho

používání přívětivé. Z důvodu své naprosté nezkušenosti s psaním takových prací mě nenapadá, jak by se schéma dalo dál rozšířit.

5 Reference

- [1] <https://en.wikipedia.org/wiki/Crossref>
- [2] <https://jasolutions.com.co/crossref/introducing-crossref-basics/>
- [3] <https://www.lib.cas.cz/pro-redakce/pridelovani-doi-crossref/>
- [4] https://cs.wikipedia.org/wiki/Digital_object_identifier
- [5] <https://support.crossref.org/hc/en-us/articles/213123586-Metadata-and-content-type-overview>
- [6] <https://www.crossref.org/get-started/>
- [7] <https://www.crossref.org/services/cited-by/>
- [8] <https://www.crossref.org/services/crossmark/>
- [9] <https://www.crossref.org/services/event-data/>
- [10] <https://www.crossref.org/services/similarity-check/>
- [11] <https://www.crossref.org/services/funder-registry/>
- [12] <http://data.crossref.org/schemas/crossref4.4.2.xsd>
- [13] <http://data.crossref.org/schemas/fundref.xsd>
- [14] <http://data.crossref.org/schemas/AccessIndicators.xsd>
- [15] <https://support.crossref.org/hc/en-us/articles/214360746>
- [16] <https://support.crossref.org/hc/en-us/articles/214572423>
- [17] <https://en.wikipedia.org/wiki/CODEN>

6 Metadata v Dublin Core

```
<dc:title>DOI/CrossRef Metadata Schema</dc:title>  
<dc:creator>Ondřej Ryšánek</dc:creator>  
<dc:date>2019-11-27</dc:date>
```

<dc:subject>CrossRef</dc:subject>

<dc:type>text</dc:type>

<dc:language>cz</dc:language>