

A szemantikus adatpublikáció gyakorlati kérdései

Tények és mítoszok

Hubay Miklós

humáninformatikus

Petőfi Irodalmi Múzeum

ELTE ITDI



Num necesse est horrere?

Z39.50

OAI

SRU/SRW

SWORD

a szemantikus web technológiája



Adatok átadása más felhasználóknak, szolgáltatásépítés céljából

Szemantikus (értelmelemmel teli) web

A jelenlegi világhálót **kiegészítő**, új réteg, amelynek alapegysége a dokumentum helyett az **adat** és a **kapcsolat**

- *Nyikolaj Szemjonovics Rubasov*
- *Winston Smith*
- *Ezerkilencszáznyolcvannégyszázötvenhárom*
- *Sötétség délben*
- *Arthur Koestler*
- *George Orwell*
- *198, [2] oldal*
- *Franklin Társulat*
- *1948*
- *963-7994-98-X*



Az Ezerkilencszáznyolcvannégyszázötvenhárom **főhőse** Winston Smith

A Sötétség délben **szerzője** Arthur Koestler

Az Ezerkilencszáznyolcvannégyszázötvenhárom **megjelenési éve** 1948

Hogy érti meg a gép?

Egységes azonosító (URI)

Tim Berners-Lee javaslata: **HTTP-formátum**

Nem dokumentumra mutat, hanem **azonosít!**

Szükség esetén azonban rendelhető hozzá további információ

<http://963-7994-98-X>

<https://viaf.org/viaf/95155403> (Orwell)

<http://www.geonames.org/719819> (Magyarország)

És a kapcsolat?

Egységes azonosító (URI)

<http://id.loc.gov/ontologies/bibframe/title>

<http://id.loc.gov/ontologies/bibframe/translationOf>

<http://id.loc.gov/ontologies/bibframe/dimensions>

Egy terület (pl. a könyvtártudomány) relációi összegyűjtve: **szótár / ontológia**

BIBFRAME / bibliotek-o / PMO / BIBO / DCRDF / ...

Kiegészíthetik egymást!

Kapcsoljuk össze!



RDF – Resource Description Framework
leíró keretrendszer

A Sötétség délben – szerzője - Arthur Koestler

Adat1 – kapcsolat – Adat2

URI – URI - URI

Három URI összetartozásának formalizálása ún. állítások formájában történik, ezeket az állításokat tripleteknek nevezzük.

Az informatikai háttér összefoglalása

Azonosítók (URI-k) a dolgoknak → **névterekből** (pl. VIAF, GeoNames, Magyar Nemzeti Névtér, Getty, stb.)

Azonosítók (URI-k) a kapcsolatoknak → **szótárakból** (pl. BIBFRAME)

Összekapcsolás az RDF segítségével → **tripletek** → adathálózat (**web of data, linked data**) → ~ **szemantikus web**

Az állítások közzététele, szolgáltatása a fogadónak megfelelő formátumban (**RDF/XML, JSON-LD, stb.**)

Miért jó a könyvtárnak?

Általános, közérthető, webes szabvány → kiszélesíti a választható informatikai megoldások körét (pl. adatvizualizáció)

Általános, közérthető, webes szabvány → az adatok konverzió után jóval szélesebb felhasználói körhöz juthatnak el, mint a MARC használatával

Új, **kiterjesztett felhasználói élményt** biztosító discovery-szolgáltatások építését teszi lehetővé

Lehetővé teszi a **funkcionális modellek** (FRBR, FRAD, FRSAD → LRM) és az ezeken alapuló **RDA** katalogizálási szabályzat a mainál jóval hatékonyabb alkalmazását

Felhasználható a **keresőbe jutáshoz**

A MARC túpírozható, de...

Az összetett entitásszerkezet és kapcsolati háló megfelelő megjelenítésére nem képes

336
337
338
264
758
\$0 - \$1



RDF – Resource Description Framework
leíró keretrendszer

URI-k beszúrása a megfelelő almezőkbe → a saját entitások összepárosítása közismert névterekkel

Mit lehet már ma?

Adatkihozatal, konverzió → szolgáltatás gráfadatbázisból (triplestore)

Integrált könyvtári platformok (Alma, OliSuite, ...)

- RDF-nézet
- Szolgáltatás API-n keresztül (választható szótár szerint)
- A teljes halmaz publikálása (választható szótár szerint)

Online metaadat-beviteli felületek (CEDAR, Sinopia, OSZK Tudástár ...)

- Tetszés szerint, több szótár alapján konfigurálható űrlapok

MARC-alapú OPAC-ok gazdagítása entitás-információkkal

- Magyar Nemzeti Múzeum Könyvtára – a MARC mellett gráfadatbázis is szolgáltat

Discovery felületek

- SHARE-VDE – entítások mentén történő keresés (művek, ágensek), névtér-kapcsolatokkal, névformákkal

Adatvizualizációs megoldások

- Petőfi Irodalmi Múzeum – a MARC-ból érkező adatok szemantikus jelölése alapján

Alma Resource management UI

Alma exposes BIBFRAME representation of a BIB record in the record simple view Alma UI:



The screenshot shows the Alma Record View interface. At the top, there is a navigation bar with a back arrow, the text "Record View", and "Back" and "Edit" buttons. Below this is a header for the record: "Data base [electronic resource]." with an information icon. A table of metadata follows, with columns for field names and values. Below the table are two tabs: "MARC" and "BIBFRAME", with "BIBFRAME" selected and highlighted with a red box. The main content area displays the XML representation of the BIBFRAME record.

MMS ID	99127876390001021	Suppress from publishing	No	Export to WorldCat	Don't publish
Export to Libraries Australia	No	Brief level	-	Record format	marc21
Originating system	SFX	Originating system ID	99954026959913	Originating system version	-

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:bf="http://id.loc.gov/ontologies/bibframe/"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:bfic="http://id.loc.gov/ontologies/bfic/" xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#">
<bf:Work rdf:about="https://open-na.hosted.exlibrisgroup.com/alma/TR_INTEGRATION_INST/99954926959913#Work">
  <bf:adminMetadata>
    <bf:AdminMetadata>
      <bf:generationProcess>
        <bf:GenerationProcess>
          <rdfs:label>DLC marc2bibframe2 v1.3.0-SNAPSHOT: 2017-09-07T05:36:11Z</rdfs:label>
        </bf:GenerationProcess>
      </bf:generationProcess>
    </bf:adminMetadata>
  </bf:Work>
  <bf:status>
    <bf:Status>
      <bf:code>n</bf:code>
    </bf:Status>
  </bf:Work>
</rdf:RDF>
```

https://developers.exlibrisgroup.com/alma/integrations/linked_data/BIBFRAME/






Enter Field Name
Date Of Work

Enter Preferred Label




Enter Field Help Text

MULTIPLE REQUIRED




- A
- Calendar
- Envelope
- #
- ...

CEDAR, Stanford




(Center for Expanded Data Annotation and Retrieval)




Place Of Origin Of The Work





Nature Of Content


Coverage Of Content

Content Type








Media Type



Find a property in BioPortal or [Enter Property URI](#)

Search in BioPortal Start Over

Content Type  

50 results for the query 'Content Type'. Click on a property below to select it

PROPERTY	DEFINITION	TYPE	SOURCE	ID
Content Type	-	Annotation Property	CARELEX	Content_Type
Content type	Categorization reflecting the fundamental form of communication in which the content is expressed and the human sense through which it is intended to be perceived.	Object Property	BIBFRAME	content
contentType	The supported content type(s) for an EntryPoint response.	Datatype Property	SCHEMA	contentType

<https://cedar.metadatascenter.org/dashboard?folderId=https:%2F%2Frepo.metadatascenter.org%2Ffolders%2Fd1f8d529-1759-4d3b-93e3-70d97c775727>

MNM Könyvtár

Relációs és gráf-adatbázis együttműködése



The screenshot shows the catalog record for the book 'A Szent Korona' by Tóth Endre. The record is displayed in a light gray box with a white background. At the top left is the logo of the Magyar Nemzeti Múzeum (Hungarian National Museum). The title 'Magyar Nemzeti Múzeum Központi Könyvtár' is at the top right. Below the title is a navigation bar with a back arrow, the text 'Vissza a találati oldalhoz', a page indicator '3./49', and a right arrow. Below the navigation bar are tabs for 'Teljes', 'Részletes', 'MARC', 'MARCXML', 'Cédula', and 'RIS'. The main content area lists the following details:

- Cím:** A Szent Korona
- Szerzőségi közlés:** Tóth Endre
- Személynév:** Tóth Endre (1944) 
- Kiadás helye:** [Budapest]
- Kiadó neve:** Kossuth Kiadó
- Megj. éve:** 2016
- Terjedelem, fizikai jellemzők:** 45 , [3] p : ill ; 28 cm
- Sorozat:** A magyar történelem rejtélyei...
- Bibliográfiai megjegyzés:** Bibliogr. : p. 44
- Nyelv:** Magyar
- ISBN:** 978-963-09-8444-7 : 795,- Ft
- Tárgyszó:** magyar történelem
Szent Korona

At the bottom left of the record is the Google Info logo. A dropdown menu is open next to the 'Személynév' field, showing the following options:

- isni.org
- viaf.org
- www.wikidata.org
- data.bnf.fr
- id.loc.gov

<https://catalog.library.hnm.hu:443/hu/record/-/record/MNMKVT248275>

EXPAND ALL CLOSE ALL

This person in



data.bnf.fr

VI AF



Orwell, George, 1903-1950

ID: 19326

Works

Other name forms

- Orwell, George, 1903-1950
- Orwell, George, pseud
- Orwell, George, 1903-1950
- Orwell, G. 1903-1950 George
- George Orwell English author and journalist
- Orwell, George

...(other forms)

Bibliography

(Click title to search on Google)

- 1984
 - Animal farm
 - Bih yād-i Kātālūniyā
 - Collected essays, journalism, and letters of George Orwell. 1968
- (other titles)

SHARE-VDE.org (Virtual Discovery Environment)

Browse



Person/Work

Italiano Info Contacts

Go to Publications

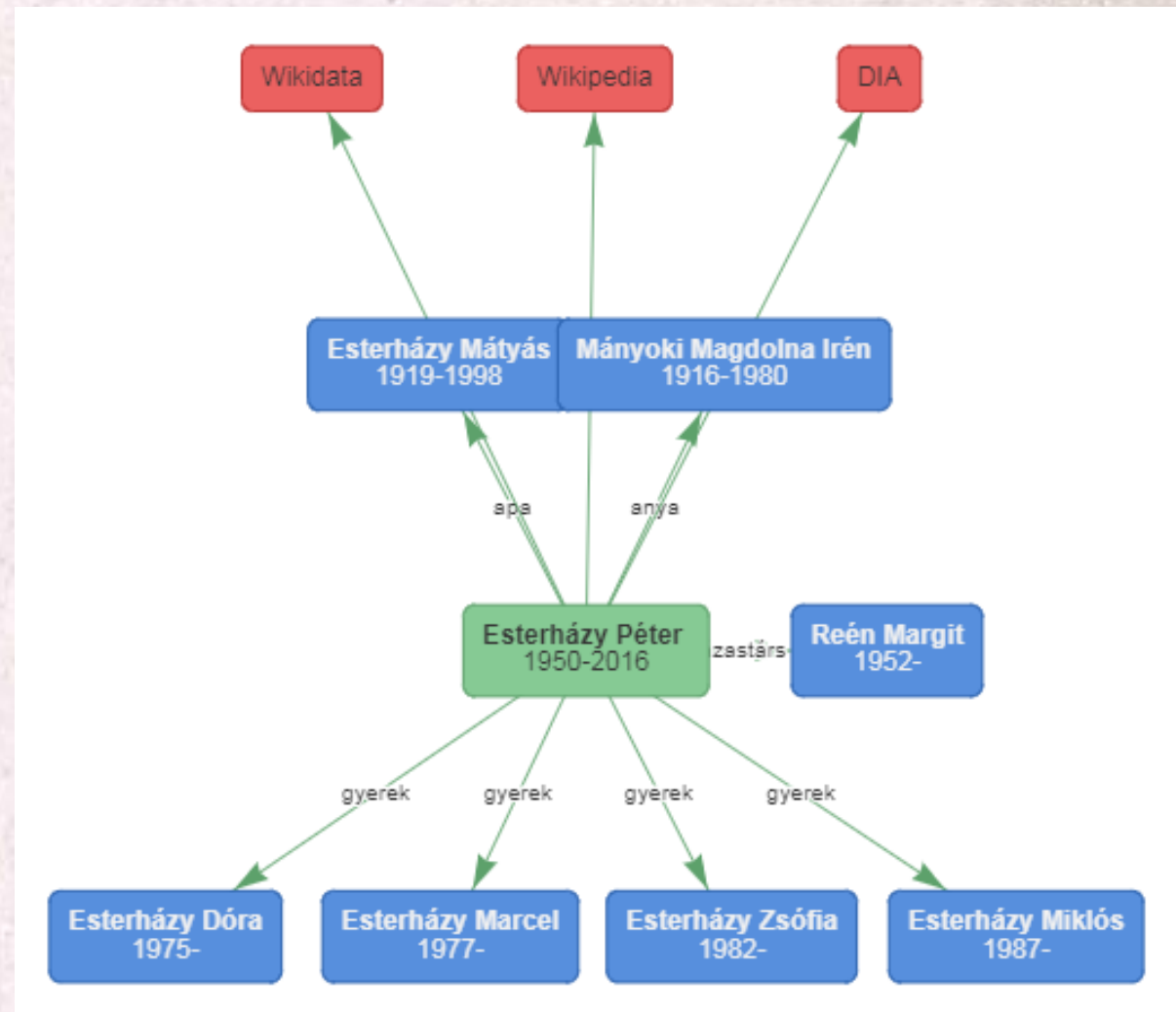
Person

Work

Publishers

Subject

A Petőfi Irodalmi Múzeum vizualizációs megoldása




<http://resolver.pim.hu/auth/PIM53067>

BIBFRAME a gyakorlatban – konverzió az Országos Széchényi Könyvtárnak




Kiindulópont, felada

- Bibliográfiai r
- MARCEdit szc
- integrált eszk
- Egyedi, testre
- felhasználásá
- Konverzió RD
- A két halmaz




MARCNext

MARCNext is a laboratory space to introduce new tools and concepts related to the integration of linked data and semantic web concepts into library metadata.




bf:
BIBFRAME


BibFrame Testbed



Link Identifiers



OpenRefine Integration



SPARQL Browser

```
Listner - [c:\Users\Hubay Miklós\Downloads\molnar_bib.mrc]
Fájl Szerkesztés Beállítások Kikódolás Súgó 15 %
01454nam a2200325 i 450099900190000001000700019003001500026005001700041008
004100058040004500099041000800144080001500152100005700167240003700224245007
400261264004500335264001100380300002800391336013000419337013600549338012300
685340001100808347001100819583003100830710008500861776015000946850001201096
852001201108942000801120. .c233151.d233151.233151.HU-IL-001206-1.201812101
60029.0.181113s1921 hu ||||| |||| 00| d hun d. .aHU-IL-001206-1.bhun.ch
U-IL-001206-1.eRDA.0 .ahun.1 .a894.511-22.1 .aMolnár Ferenc,.d1878-1952.es
zerzL'.4aut.04016.94016.12.060194.aA hattyaş..lmagyar.960194.12.aA hattyaş
:.bvÅ-gjÅ~tÅ@k hÅ~rom felvonÅ~sban /.cÅ~rta MolnÅ~r Ferenc.. 1.aBudapest :.
bFranklin-TÅ~rsulat. .c[1921]. 4.cÅ@1921. .a198, [2] oldal ;.c14 cm. .2rd
acontent.aszÅ~veg.btxt.0http://rdaregistry.info/termList/RDAContentType/102
0.0http://id.loc.gov/vocabulary/contentTypes/txt. .2rdamedia.asegÅ@deszkÅ~
z nÅ@lkÅ@li.bn.0http://rdaregistry.info/termList/RDAMediaType/1007.0http://
id.loc.gov/vocabulary/mediaTypes/n. .2rdacarrier.akÅ~tet.bnc.0http://rdare
gistry.info/termList/RDACarrierType/1049.0http://id.loc.gov/vocabulary/carr
iers/nc. .2rdagen. .2rdaft. .xszemantikus publikÅ~lÅ~sra.2 .aFranklin T
Å~rsulat Magyar Irodalmi IntÅ@zet Å@s KÅ~nyvnyomda.ekiadÅ~.4pbl.960166.08.i
Elektronikus nÅ~solata.aMolnÅ~r Ferenc, 1878-1952.dBudapest : [OrszÅ~gos Sz
Å@chÅ@nyi KÅ~nyvtÅ~r DigitalizÅ~lÅ~ OsztÅ~ly], 2018.tA hattyaş.w233166. .a
HuBpOSK. .aHuBpOSK. .2ddc..01677nam a2200361 i 4500999001900000010007000
190030015000260050017000410080041000580400045000990410008001440800015001521
000057001672400037002242450064002612640045003252640011003703000028003813360
130004093370136005393380123006753400011007983470011008095050154008205830031
009747000077010057000057010827000059011397100085011988500012012838520012012
95942000801307. .c233152.d233152.233152.HU-IL-001206-1.20181210160109.0.18
1113s1921 hu ||||| |||| 00| d hun d. .aHU-IL-001206-1.bhun.chU-IL-00120
6-1.eRDA.0 .ahun.1 .a894.511-25.1 .aMolnÅ~r Ferenc,.d1878-1952.eszerzL'.4au
t.04016.94016.1 .060207.aSzÅ~nhÅ~z..lmagyar.960207.10.aSzÅ~nhÅ~z :.bhÅ~rom
egyfelvonÅ~sos /.cÅ~rta MolnÅ~r Ferenc.. 1.aBudapest :.bFranklin-TÅ~rsulat,
.c[1921]. 4.cÅ@1921. .a215, [1] oldal ;.c14 cm. .2rdacontent.aszÅ~veg.bt
xt.0http://rdaregistry.info/termList/RDAContentType/1020.0http://id.loc.gov
/vocabulary/contentTypes/txt. .2rdamedia.asegÅ@deszkÅ~z nÅ@lkÅ@li.bn.0http
://rdaregistry.info/termList/RDAMediaType/1007.0http://id.loc.gov/vocabular
y/mediaTypes/n. .2rdacarrier.akÅ~tet.bnc.0http://rdaregistry.info/termList
/RDACarrierType/1049.0http://id.loc.gov/vocabulary/carriers/nc. .2rdagen.
.2rdaft. .00.gI..tEIL'jÅ~tÅ@k Lear kirÅ~lyhoz : szatÅ~ra egy felvonÅ~sban.g
II..tMarsall : szÅ~nmL± egy felvonÅ~sban.gIII..tÅz ibolya : vÅ-gjÅ~tÅ@k egy
felvonÅ~sban. .xszemantikus publikÅ~lÅ~sra.12.aMolnÅ~r Ferenc,.d1878-1952
..tEIL'jÅ~tÅ@k Lear kirÅ~lyhoz..lmagyar.960202.12.aMolnÅ~r Ferenc,.d1878-19
52..tMarsall..lmagyar.960204.12.aMolnÅ~r Ferenc,.d1878-1952..tÅz ibolya..lm
agyar.960199.2 .aFranklin TÅ~rsulat Magyar Irodalmi IntÅ@zet Å@s KÅ~nyvnyom
da.ekiadÅ~.4pbl.960166. .aHuBpOSK. .aHuBpOSK. .2ddc..01304nam a2200313 i
4500999001900000010007000190030008000260050017000340080041000510400045000
920410008001370800015001451000057001602400043002172450087002602640045003472
```

</bf:hasInstance>

Redundancia

- **337 (médiatípus, RDA)**
- és **007/00** (anyag kategóriája)
- és **000/06** (rekordtípus)

007--ELECTRONIC RESOURCE	
00 - Category of material	
c - Computer file	I - Electronic (if not already added from Leader) Add I - media - Media - http://id.loc.gov/vocabulary/mediaTypes/c (if not added from 337 Media Type); Add rdfs:label "computer"
01 - Specific material designation	I - carrier - Carrier ; Add these if not already added from 338 Carrier Type

- **338 (hordozótípus, RDA)**
- és **007/01** (specifikus anyagmegjelölés)
- és **008/23** (példányforma (*item form*))

Részleges redundancia

```
<bf:descriptionConventions>
```

```
<bf:DescriptionConventions
```

```
rdf:about="http://id.loc.gov/vocabulary/descriptionConventions/040$e értéke" />
```

```
</bf:descriptionConventions>
```

```
<bf:descriptionConventions>
```

```
<bf:DescriptionConventions>
```

```
<bf:code>000/18 alapján</bf:code>
```

```
</bf:DescriptionConventions>
```

```
</bf:descriptionConventions>
```

**Inkább
központozási
adatok**

Művek tulajdonságai a rögzített hosszúságú mezők alapján

```
<rdf:type rdf:resource=  
"http://id.loc.gov/ontologies/bibframe/000/06 alapján" />
```

```
<bf:descriptionConventions>  
<bf:DescriptionConventions>  
<bf:code>000/18 alapján</bf:code>  
</bf:DescriptionConventions>  
</bf:descriptionConventions>
```

```
<bf:status>  
<bf:Status>  
<bf:code>000/05</bf:code>  
</bf:Status>  
</bf:status>
```

```
<bflc:encodingLevel>  
<bflc:EncodingLevel>  
<bf:code>000/17</bf:code>  
</bflc:EncodingLevel>  
</bflc:encodingLevel>
```

```
<bf:genreForm>  
<bf:GenreForm rdf:about=  
"http://id.loc.gov/vocabulary/marcgt/008/33 alapján">  
<rdfs:label>008/33 alapján</rdfs:label>  
</bf:GenreForm>  
</bf:genreForm>
```

```
<bf:language>  
<bf:Language rdf:about=  
"http://id.loc.gov/vocabulary/languages/008/35-37" />  
</bf:language>
```

```
<bf:colorContent>  
<bf:ColorContent rdf:about=  
"http://id.loc.gov/vocabulary/mcolor/007/03 alapján">  
<rdfs:label>007/03 alapján</rdfs:label>  
</bf:ColorContent>  
</bf:colorContent>
```

Megjelenési formák tulajdonságai a rögzített hosszúságú mezők alapján

```
<bf:issuance>  
<bf:Issuance rdf:about=  
"http://id.loc.gov/vocabulary/issuance/000/07 alapján" />  
</bf:issuance>
```

```
<bf:digitalCharacteristic>  
<bf:DigitalCharacteristic>  
<rdf:type rdf:resource=  
"http://id.loc.gov/ontologies/bibframe/EncodingFormat" />  
<rdfs:label>007/09 alapján</rdfs:label>  
</bf:DigitalCharacteristic>  
</bf:digitalCharacteristic>
```

```
<bf:digitalCharacteristic>  
<bf:DigitalCharacteristic>  
<rdf:type rdf:resource=  
"http://id.loc.gov/ontologies/bflc/ImageBitDepth" />  
<rdfs:label>007/06-08</rdfs:label>  
</bf:DigitalCharacteristic>  
</bf:digitalCharacteristic>
```

```
<bf:provisionActivity>  
<bf:ProvisionActivity>  
<rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Publication" />  
<bf:date rdf:datatype=  
"http://id.loc.gov/datatypes/edtf">008/07-10</bf:date>  
<bf:place>  
<bf:Place rdf:about=  
"http://id.loc.gov/vocabulary/countries/008/15-17" />  
</bf:place>  
</bf:ProvisionActivity>  
</bf:provisionActivity>
```

```
<rdf:type rdf:resource=  
"http://id.loc.gov/ontologies/bibframe/007/00 alapján" />
```

```
<bf:illustrativeContent>  
<bf:Illustration rdf:about=  
"http://id.loc.gov/vocabulary/millus/008/18-21">  
<rdfs:label>008/18-21 alapján</rdfs:label>  
</bf:Illustration>  
</bf:illustrativeContent>
```

MarcKey, MatchKey, SortKey, Label

- A BFLC „hivatalos” kiegészítésében definiált tulajdonságok a személyek/testületek, valamint címek megfelelő konverziójához
- Főszabály: az almezők sorrendjének megtartása, majd...
- **MatchKey** – indikátorértékek és almezőjelek eldobása
- **MarcKey** – indikátorértékek és almezőjelek megtartása
- **rdfs:label** – indikátorértékek és almezőjelek eldobása
- **titleSortKey** – rdfs:label + a második indikátorban jelzett kihagyandó karakterek eldobása

*name00MarcKey
name00MatchKey
name10MarcKey
name10MatchKey
name11MarcKey
name11MatchKey*

*primaryContributorName00MatchKey
primaryContributorName10MatchKey
primaryContributorName11MatchKey*

*title00MarcKey
title00MatchKey
title10MarcKey
title10MatchKey
title11MarcKey
title11MatchKey
title30MarcKey
title30MatchKey
title40MarcKey
title40MatchKey*

titleSortKey

A \$0 értékek problémája

=710 2\

\$a Franklin Társulat Magyar
Irodalmi Intézet és

Könyvnyomda

\$e kiadó

\$4 pbl

\$9 60166

*Nem szerepeltek \$0 értékek
(azonosítók) a téves űrlap-
beállítás miatt*

```
<bf:contribution>
```

```
<bf:Contribution>
```

```
<bf:agent>
```

```
<bf:Agent rdf:about="http://data.koha.oszk.hu/233151#Agent">
```

```
<rdf:type
```

```
  rdf:resource="http://id.loc.gov/ontologies/bibframe/Organization"/>
```

```
<bflc:name10MatchKey>Franklin Társulat Magyar Irodalmi Intézet és  
Könyvnyomda</bflc:name10MatchKey>
```

```
<bflc:name10MarcKey>7102 $aFranklin Társulat Magyar Irodalmi Intézet és  
Könyvnyomda$ekiadó$4pbl$960166</bflc:name10MarcKey>
```

```
<rdfs:label>Franklin Társulat Magyar Irodalmi Intézet és  
Könyvnyomda</rdfs:label>
```

```
</bf:Agent>
```

```
</bf:agent>
```

A Franklin Társulat több
azonosítót kapott! ☹️



A keresőbe jutás mítosza



<https://dataliberate.com/2018/08/27/bibframe-schema-org-chocolate-teapots/>



A keresőbe jutás mítosza

- Az oldalaknak állandó URL-je van (linkfeloldás)
- Oldaltérkép / robots.txt segíti az indexelőrobot munkáját
- Az oldalakra linkek mutatnak (linkcsere pl. Wikidatával, Wikipediával)

schema.org

- általános, a keresők által is használt szótár, könyvtári dokumentumokra is alkalmazható
- az OPAC megjelenítő sablonjában kell elhelyezni
- több discovery szoftver már tudja automatikusan

<https://kifu.videotorium.hu/hu/recordings/24689/online-katalogusok-felhozasa-a-felszini-webre>

Köszönöm a figyelmet!



Bővebben:

Hubay Miklós

Tények, mítoszok és lehetőségek a szemantikus web világában

Könyvtári Figyelő 2019/2 (megjelenés alatt)