ExLibris
Part of **Clarivate**

Extracting Linked Data from Alma via the general publishing profile

Yoel Kortick   **May 2024**

Agenda

- Introduction
- The sample bibliographic record
- The publishing profile
- The output with Linked Data

Introduction

ExLibris
Part of **Clarivate**

Introduction

- The <u>General Publishing Profile</u> in Alma allows the institutions to publish (export data) in various formats to third-party systems, and to enrich the data as it is exported.

- The data can be both bibliographic as well as authority and can be enriched in various ways when published.

- This presentation will focus on publishing bibliographic data and enriching the data with Linked Data when it is published.

Introduction

- When Linked Data is added to the published record it is added in the form of a Uniform Resource Identifier (URI).

- The URI is added for each heading which is linked to an authority record. The headings which include linked data must be linked to an authority record.

- These URIs are an essential part of the linked data that might be used to retrieve new information from related entities.

- In a Linked Data context, the user would follow the links and "crawl" from one data description to another.

- For an extensive list of authority vocabularies for which Alma adds a URI see the "Enrichment" section at Linked Data.

Introduction

- In this presentation we will first look at a sample record which has many headings linked to a wide variety of authority vocabularies.

- We will then publish the record using the General Publishing profile with Linked Data Enrichment.

- We will then look at the output file and see the URIs added to the subfield 0 of the relevant fields.

- Finally, we will see what data those URIs bring by pasting them into the URL search bar of a browser.

- The data brought by the URIs could be used in an application taking advantage of the Linked Open Data.

Introduction

- In addition to this presentation see also:
  - Creating a General Publishing Profile
  - Linked Data
  - Working with Authority Records

The sample bibliographic record

ExLibris
Part of **Clarivate**

The sample bibliographic record

- To see how headings from a bibliographic record get enriched with Linked data during publishing, we will look at one specific record and see what happens with the headings.

- The sample record which we will look at includes headings which are linked to a wide variety of authority vocabularies.

- The title of the record is "Homo Deus : techno-humanism vs dataism and the place of information in society : lectio magistralis in Library science / by Philip E. Schreur."

The sample bibliographic record

- Here is a 100 field and 600 field linked to the LCNAMES (Library of Congress Names) authority vocabulary

082	1 4	$$a 021 $$2 14
🔍 100	1	$$a Schreur, Philip Evan, $$e author.
245	1 0	$$a Homo Deus : $$b techno-humanism vs dataism and the place of information in society : lectio magistralis in Library science / $$c by Philip E. Schreur.
250		$$a Prima edizione.
264	1	$$a Fiesole (Firenze) : $$b Casalini Libri, $$c 2023.
300		$$a 51 pages ; $$c 24 cm.
490	0	$$a Letture magistrali in biblioteconomia ; $$v XVI
500		$$a Lecture held in Florence, Italy, March 7, 2023.
546		$$a Text in English with Italian translation; abstract in English and Italian.
🔍 650	0	$$a Information science.
🔍 650	0	$$a Information technology.
🔍 650	0	$$a Metadata.
🔍 650	0	$$a Information society.
🔍 600	1 0	$$a Harari, Yuval N. $$t Historyah shel ha-mahar.
🔍 650	2	$$a Information Technology.

The sample bibliographic record

- Here are 650_0 fields linked to the LCSH (Library of Congress Subject Headings) authority vocabulary

	546		$$a Text in English with Italian translation; abstract in English and Italian.
🔭	650	0	$$a Information science.
🔭	650	0	$$a Information technology.
🔭	650	0	$$a Metadata.
🔭	650	0	$$a Information society.
🔭	600 1	0	$$a Harari, Yuval N. $$t Historyah shel ha-maḥar.

The sample bibliographic record

- Here is a 650_2 field linked to the MESH (Medical Subject Headings) authority vocabulary

🔭	600 1	0	$$a Harari, Yuval N. $$t Historyah shel ha-maḥar.
🔭	650	2	$$a Information Technology.
🔭	650	7	$$a Bibliotecas $$x Tecnología de la información. $$2 embne

ExLibris
Part of Clarivate

The sample bibliographic record

- Here is a 650_7 field with subfield 2 "embne" linked to the National Library of Spain (SpMaBN) authority vocabulary. This was added in Nov. 2023.

	650	2	$$a Information Technology.
	650	7	$$a Bibliotecas $$x Tecnología de la información. $$2 embne
	650	7	$$a Information technology. $$9 eng $$2 eurovoc

The sample bibliographic record

- Here is a 650_7 field with subfield 2 "eurovoc" linked to the European Union Vocabularies vocabulary. This was added in Aug. 2023.

650	7	$$a Bibliotecas $$x Tecnología de la información. $$2 embne
650	7	$$a Information technology. $$9 eng $$2 eurovoc
650	7	$$a Bibliothèques techniques. $$2 idref

The sample bibliographic record

- Here are 650_7 fields with subfield 2 "idref" linked to the IdRef – Identifiants et référentiels pour l'Enseignement supérieur et la Recherche vocabulary. This was added in Feb. 2023.

	650	7	$$a Information technology. $$9 eng $$2 eurovoc
	650	7	$$a Bibliothèques techniques. $$2 idref
	650	7	$$a Informatique. $$2 idref
	650	7	$$a Informationstechnik. $$2 gnd

ExLibris
Part of **Clarivate**

The sample bibliographic record

- Here is a 650_7 field with subfield 2 "gnd" linked to the Gemeinsame Normdatei (Integrated Authority File) of the Deutsche Nationalbibliothek vocabulary

	650	7	$$a Informatique. $$2 idref
	650	7	$$a Informationstechnik. $$2 gnd
	650	7	$$a Information science. $$9 lat $$2 nli

The sample bibliographic record

- Here is a 650_7 field with subfield 2 "nli" linked to the National Library of Israel vocabulary

	650	7	$$a Informationstechnik. $$2 gnd
	650	7	$$a Information science. $$9 lat $$2 nli
	650	7	$$a Information science $$2 fast

The sample bibliographic record

- Here are 650_7 fields with subfield 2 "fast" linked to the Faceted Application of Subject Terminology vocabulary

The sample bibliographic record

- Here is a 655_7 field with subfield 2 "lcgft" linked to the Library of Congress Genre/Form Terms vocabulary

🔍	650	7	$$a Metadata $$2 fast
🔍	655	7	$$a Academic theses. $$2 lcgft
	700 1	2	$$a Schreur, Philip Evan. $$b Homo Deus.

The sample bibliographic record

🔭	100	1	$$a Schreur, Philip Evan, $$e author.
	245	1 0	$$a Homo Deus : $$b techno-humanism vs dataism and the place of information in society : lectio magistralis in Library science / $$c by Philip E. Schreur.
	250		$$a Prima edizione.
	264	1	$$a Fiesole (Firenze) : $$b Casalini Libri, $$c 2023.
	300		$$a 51 pages ; $$c 24 cm.
	490	0	$$a Letture magistrali in biblioteconomia ; $$v XVI
	500		$$a Lecture held in Florence, Italy, March 7, 2023.
	546		$$a Text in English with Italian translation; abstract in English and Italian.
🔭	650	0	$$a Information science.
🔭	650	0	$$a Information technology.
🔭	650	0	$$a Metadata.
🔭	650	0	$$a Information society.
🔭	600	1 0	$$a Harari, Yuval N. $$t Historyah shel ha-maḥar.
🔭	650	2	$$a Information Technology.
🔭	650	7	$$a Bibliotecas $$x Tecnología de la información. $$2 embne
🔭	650	7	$$a Information technology. $$9 eng $$2 eurovoc
🔭	650	7	$$a Bibliothèques techniques. $$2 idref
🔭	650	7	$$a Informatique. $$2 idref
🔭	650	7	$$a Informationstechnik. $$2 gnd
🔭	650	7	$$a Information science. $$9 lat $$2 nli
🔭	650	7	$$a Information science $$2 fast
🔭	650	7	$$a Information society $$2 fast
🔭	650	7	$$a Information technology $$2 fast
🔭	650	7	$$a Metadata $$2 fast
🔭	655	7	$$a Academic theses. $$2 lcgft

- LCNAMES
- LCSH
- MESH
- National Library of Spain (SpMaBN)
- European Union Vocabularies
- IdRef – Identifiants et référentiels pour l'Enseignement supérieur et la Recherche
- Deutsche Nationalbibliothek
- National Library of Israel
- Library of Congress Genre/Form Terms vocabulary
- Faceted Application of Subject Terminology

The publishing profile

ExLibris.
Part of **Clarivate**

The publishing profile

- As previously stated, the <u>General Publishing Profile</u> in Alma allows the institutions to publish (export data) in various formats to third-party systems, and to enrich the data as it is exported.

- To include linked data in the published records the checkbox "Linked Data Enrichment" should be checked in the "Data Enrichment" tab.

- To publish the records in MARC21 format with the linked data check the "Linked Data Enrichment" box in the "Data Enrichment" tab.

  - This will add the $$0 (subfield zero) with linked data URIs to certain records when publishing. This is relevant only when the output format is MARC. When the output format is RDA/RDF the Linked data enrichment check box is automatically selected.

ExLibris
Part of **Clarivate**

The publishing profile

Profile Details

Content Type **Bibliographic**

Profile Name * **Linked Data Enrichment**

Profile Description

On the "Profile Details" tab we have named the publishing profile "Linked Data Enrichment"

Publishing Parameters

Status ● Active ○ Inactive

Scheduling **Not scheduled** Edit Scheduling

Email Notifications

Content

Set name * homo deus techno humanism

Additional set name Select from a list

Publish the entire repository ☐

Filter out the data using

Publish on: **Bibliographic level**

Output format **MARC21 Bibliographic**

We will be working on a set of one record which includes the title we showed here in the "The sample bibliographic record" section

We have chosen output format "MARC21 Bibliographic"

The publishing profile

Publishing Protocol

FTP ☑

FTP configuration *   Ex Libris Secure FTP Service

Sub-directory   **Linked Data Enrichment**

Include all records each time file is published ☑

Disable file compression ☐

Compressed file extension   tar.gz

Physical format   XML

Number of records in file   1000

Use default file name   ○ No   ● Yes
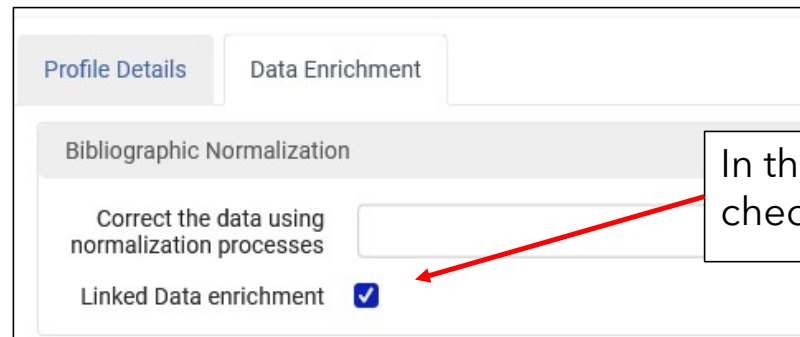
File name prefix *   **Linked_Data_Enrichment**

We will be publishing to FTP to a subdirectory "Linked Data Enrichment"

We have chosen physical format XML

The file name prefix will be Linked_Data_Enrichment

The publishing profile

Profile Details	Data Enrichment

Bibliographic Normalization

Correct the data using
normalization processes

Linked Data enrichment ☑

In the "Data Enrichment" tab we have
checked "Linked Data enrichment"

The output with Linked Data

The output with Linked Data

- The <u>General Publishing Profile</u> has run, and the output file is on the FTP server in the specified subdirectory with the specified file name prefix

As specified in the publishing profile we published to FTP to a subdirectory "Linked Data Enrichment"

Remote site:	/production/Linked Data Enrichment				
⊟ ▮ /					
⊟ ▮ production					

Filename	Filesize	Filetype	Last modified	Permis...	Owner/...
..					
Linked_Data_Enrichment_2024033115_7702909930000121_ne...	1,456	GZ File	3/31/2024 6:16:08 PM	-rwxrw...	0 0

As specified in the publishing profile the file name prefix is Linked_Data_Enrichment

The output with Linked Data

- In the output file we can see that the URIs have been added to the relevant fields in subfield 0. Here are two URIs for the 100 field.

```
<datafield tag="100" ind1="1" ind2=" ">
    <subfield code="a">Schreur, Philip Evan,</subfield>
    <subfield code="e">author.</subfield>
    <subfield code="0">http://id.loc.gov/authorities/names/n89628077</subfield>
    <subfield code="0">http://viaf.org/viaf/sourceID/LC|n89628077</subfield>
</datafield>
<datafield tag="245" ind1="1" ind2="0">
    <subfield code="a">Homo Deus :</subfield>
    <subfield code="b">techno-humanism vs dataism and the place of information in society : lectio magistralis in Library science /
    </subfield>
    <subfield code="c">by Philip E. Schreur.</subfield>
</datafield>
```

The output with Linked Data

- Here is (part of) the first URI http://id.loc.gov/authorities/names/n89628077 for the 100 field <subfield code="a">Schreur, Philip Evan,</subfield>

The output with Linked Data

- Here is (part of) the second URI http://viaf.org/viaf/sourceID/LC|n89628077 for the 100 field <subfield code="a">Schreur, Philip Evan,</subfield>

100 field linked to the LCNAMES (Library of Congress Names) authority vocabulary

The output with Linked Data

```xml
<datafield tag="650" ind1=" " ind2="2">
    <subfield code="a">Information Technology.</subfield>
    <subfield code="0">http://id.nlm.nih.gov/mesh/D000073256</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2="7">
    <subfield code="a">Bibliotecas</subfield>
    <subfield code="x">Tecnología de la información.</subfield>
    <subfield code="2">embne</subfield>
    <subfield code="0">https://datos.bne.es/tema/XX775438</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2="7">
    <subfield code="a">Information technology.</subfield>
    <subfield code="9">eng</subfield>
    <subfield code="2">eurovoc</subfield>
    <subfield code="0">http://eurovoc.europa.eu/5188</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2="7">
    <subfield code="a">Bibliothèques techniques.</subfield>
    <subfield code="2">idref</subfield>
    <subfield code="0">https://www.idref.fr/033543054</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2="7">
    <subfield code="a">Informatique.</subfield>
    <subfield code="2">idref</subfield>
    <subfield code="0">https://www.idref.fr/114596476</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2="7">
    <subfield code="a">Informationstechnik.</subfield>
    <subfield code="2">gnd</subfield>
    <subfield code="0">http://d-nb.info/gnd/040269264</subfield>
</datafield>
```

- Here are some of the URIs for the 650 fields

The output with Linked Data

```
<datafield tag="650" ind1=" " ind2="7">
    <subfield code="a">Information science.</subfield>
    <subfield code="9">lat</subfield>
    <subfield code="2">nli</subfield>
    <subfield code="0">https://open-eu.hosted.exlibrisgroup.com/alma/972NNL_INST/authorities/987007550614705171.jsonld</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2="7">
    <subfield code="a">Information science</subfield>
    <subfield code="2">fast</subfield>
    <subfield code="0">http://id.worldcat.org/fast/972640</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2="7">
    <subfield code="a">Information society</subfield>
    <subfield code="2">fast</subfield>
    <subfield code="0">http://id.worldcat.org/fast/972767</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2="7">
    <subfield code="a">Information technology</subfield>
    <subfield code="2">fast</subfield>
    <subfield code="0">http://id.worldcat.org/fast/973089</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2="7">
    <subfield code="a">Metadata</subfield>
    <subfield code="2">fast</subfield>
    <subfield code="0">http://id.worldcat.org/fast/1017519</subfield>
</datafield>
<datafield tag="655" ind1=" " ind2="7">
    <subfield code="a">Academic theses.</subfield>
    <subfield code="2">lcgft</subfield>
    <subfield code="0">http://id.loc.gov/authorities/genreForms/gf2014026039</subfield>
</datafield>
```

- Here are some more of the URIs for the 650 fields, as well as the URI for the 655 field

The output with Linked Data

- Here is part of the URI http://id.nlm.nih.gov/mesh/D000073256 for the 650 field <subfield code="a">Information Technology.</subfield>

Medical Subject Headings RDF

Home   SPARQL Query Editor   Lookup   Documentation   Sample Queries   Download   API   MeSH Home

About: Information Technology
http://id.nlm.nih.gov/mesh/D000073256

650_2 field linked to the MESH (Medical Subject Headings) authority vocabulary

Type: MeSH TopicalDescriptor
Topical Descriptors indicate the subject of an indexed item such as a journal article. See D063926, 'Drug Hypersensitivity Syndrome', for an example.
more types...

The output with Linked Data

- Here is part of the URI https://datos.bne.es/tema/XX775438 for the 650 field <subfield code="a">Bibliotecas</subfield>

> **Bibliotecas--Tecnología de la información**
>
> | Término aceptado | [es] Bibliotecas--Tecnología de la información |
> |---|---|
> | semanticRelation | http://datos.bne.es/resource/XX524917 |
> | | http://datos.bne.es/resource/XX4357184 |

650_7 field with subfield 2 "embne" linked to the National Library of Spain (SpMaBN) authority vocabulary

The output with Linked Data

- Here is part of the URI http://eurovoc.europa.eu/5188 for the 650 field
  <subfield code="a">Information technology.</subfield>

```
-<rdf:RDF>
 -<rdf:Description rdf:about="http://eurovoc.europa.eu/3636">
    <skos:related rdf:resource="http://eurovoc.europa.eu/5188"/>
 </rdf:Description>
 -<rdf:Description rdf:about="http://eurovoc.europa.eu/1376">
    <skos:related rdf:resource="http://eurovoc.europa.eu/5188"/>
 </rdf:Description>
 -<rdf:Description rdf:about="http://eurovoc.europa.eu/5188">
    <rdf:type rdf:resource="http://eurovoc.europa.eu/schema#ThesaurusConcept"/>
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1995-10-02</dcterms:created>
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2018-08-07</dcterms:modified>
    <owl:versionInfo>n/a</owl:versionInfo>
    <skos:inScheme rdf:resource="http://eurovoc.europa.eu/100141"/>
    <skos:inScheme rdf:resource="http://eurovoc.europa.eu/100221"/>
```

650_7 field with subfield 2 "eurovoc" linked to the European Union Vocabularies vocabulary

The output with Linked Data

- Here is part of the URI https://www.idref.fr/033543054 for the 650 field
  <subfield code="a">Bibliothèques techniques.</subfield>

Identifiant pérenne de la notice : 033543054

URI de l'entité : http://www.idref.fr/033543054/id

Notice de type Rameau

Point d'accès autorisé

Bibliothèques techniques

> 650_7 fields with subfield 2 "idref" linked to the IdRef – Identifiants et référentiels pour l'Enseignement supérieur et la Recherche vocabulary

Variantes de point d'accès

Bibliothèques technologiques   [Nom commun]
Technique -- Bibliothèques   [Nom commun]
Technologie -- Bibliothèques   [Nom commun]

ExLibris
Part of **Clarivate**

The output with Linked Data

- Here is part of the URI https://www.idref.fr/114596476 for the 650 field <subfield code="a">Informatique.</subfield>

Identifiant pérenne de la notice : 114596476

URI de l'entité : http://www.idref.fr/114596476/id

Notice de type FMeSH

Point d'accès autorisé

Informatique

650_7 fields with subfield 2 "idref" linked to the IdRef – Identifiants et référentiels pour l'Enseignement supérieur et la Recherche vocabulary

Informations

Note publique d''information :
The field of information science concerned with the analysis and dissemination of data through the application of computers

The output with Linked Data

• Here is part of the URI
https://open-eu.hosted.exlibrisgroup.com/alma/972NNL_INST/authorities/987007550614705171.jsonld
for the 650 field <subfield code="a">Information science.</subfield>

```
▼ subject:
    preferred_label:     "علم المعلومات"
  ▼ see_also:
      0:                 "Information literacy"
      1:                 "Communication"
▼ @id:                   "https://open-eu.hosted.exlibrisgroup.com/alma/972NNL_INST/authorities/987007550614705171"
▼ source:                "LC database, Feb. 20, 1998 (informatics 15 times more likely on a computer science book than on an information science book)"
    title:               "علم المعلومات"
▼ @context:              "https://open-eu.hosted.exlibrisgroup.com/alma/contexts/authority"
```

650_7 field with subfield 2 "nli" linked to the National Library of Israel vocabulary

The output with Linked Data

- Here is part of the URI http://id.worldcat.org/fast/972640 for the 650 field <subfield code="a">Information science</subfield>
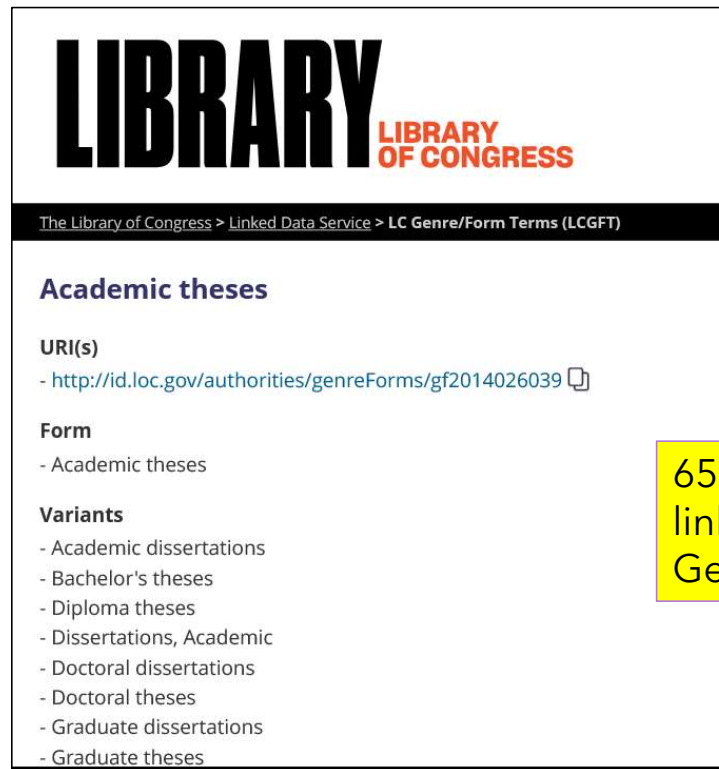
650_7 fields with subfield 2 "fast" linked to the Faceted Application of Subject Terminology vocabulary

The output with Linked Data

- Here is part of the URI http://id.loc.gov/authorities/genreForms/gf2014026039 for the 655 field <subfield code="a">Academic theses.</subfield>

655_7 field with subfield 2 "lcgft" linked to the Library of Congress Genre/Form Terms vocabulary

ExLibris

Part of **Clarivate**