

# Veröffentlichung der HeBIS Titeldaten als Linked-Data

von Dr. Thomas Striffler

## Einführung

Die Veröffentlichung bibliothekarischer Daten als Linked-Data ist für Bibliotheken aus mehrerlei Gründen, die hier nachfolgend aufgelistet werden, von zunehmender Bedeutung:

- Erhöhung der eigenen Sichtbarkeit im WWW
- Verbesserung der Auffindbarkeit der eigenen Bestände im WWW
- Sicherstellung der Nutzbarkeit bibliothekarischer Daten auch durch nichtbibliothekarische Akteure
- Anreicherung der Informationen aus Bibliotheks- und Verbunddatenbanken durch Verknüpfung mit anderen Informationen
- Verbesserung der Recherchemöglichkeiten

Siehe:

[http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl\\_ostrowski\\_2010\\_linked-data.pdf](http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl_ostrowski_2010_linked-data.pdf)

Um dieser Entwicklung gerecht zu werden, wird auch innerhalb des HeBIS-Verbundes an einer Veröffentlichung der Titeldaten als Linked-Data gearbeitet. Schon jetzt kann man sich die Daten eines einzelnen Titels im RDF/XML-Format herunterladen oder sich diese innerhalb des Browsers anzeigen lassen, so fern der verwendete Browser hierzu in der Lage ist,

## Abfrage von HeBIS-Titeldaten im RDF/XML-Format

Beispielsweise können die RDF/XML-Daten zum Titel mit der IPN 28068667 über folgende URL erreicht werden: <http://resolver.hebis.de/catalog/rdf/28068667>. In gleicher Weise können so auch die RDF/XML-Daten eines jeden anderen Titels des HeBIS-Verbundes angesprochen werden. Der Server [resolver.hebis.de](http://resolver.hebis.de) ist aber auch in der Lage ein Content Negotiation durchzuführen. So wird der Nutzer beim Aufruf der URL <http://resolver.hebis.de/catalog/28068667> auf die zu diesem Titel gehörende Seite des Verbund-OPACs weitergeleitet. Enthält diese Anfrage eines Clients im Header jedoch noch den Zusatz "Accept: application/rdf+xml", so liefert der Server auf diese Anfrage wieder die Daten im RDF/XML-Format an den Client.

Die Ressource aus dem obigen Beispiel, die mittels der RDF/XML-Daten beschrieben wird, besitzt die URI: <http://resolver.hebis.de/resource/28068667>. Diese URI ist das Subjekt in den Tripeln, mit denen die Ressource beschrieben werden kann.

## Funktionsweise der RDF/XML-Datenerzeugung

Bei einer Anfrage nach RDF/XML-Daten konvertiert der Server die aus der Datenbank kommenden Daten vom Pica+-Format in das RDF/XML-Format. Dies

geschieht durch ein über CGI aufgerufenes C-Programm (namens "rdf"). Die Art und Weise der Datenkonvertierung ist hierbei vollständig in einer FCV-Tabelle hinterlegt, auf die das zuvor erwähnte C-Programm bei der Konvertierung zugreift. Eine Änderung des verwendeten RDF-Vokabulars kann somit allein über eine Bearbeitung der FCV-Tabelle erfolgen.

### Konvertierung der Daten von Pica+ nach RDF/XML

Bei der Konvertierung werden die einzelnen Subfelder der Kategorien eines Pica+-Datensatzes in Prädikate der Tripel gewandelt, die die Ressource beschreiben. Wo es möglich ist, werden die Subfelder mit URIs aus bereits bekannten Ontologien gemappt. So wird beispielsweise das Subfeld 033A \$n, welches die Verlagsangabe enthält, auf das Dublin Core Element `dcterms:publisher` (<http://purl.org/dc/elements/1.1/publisher>) gemappt. Zur Beschreibung der Daten im RDF/XML-Format wurde das gleiche Vokabular verwendet, das auch vom HBZ und der Universitätsbibliothek Mannheim verwendet wird. Einen Überblick über das Vokabular, das aus den Ontologien Dublin Core (`dcterms`), Bibliographic Ontology (`bibo`), Web Ontology Language (`owl`) und International Standard Bibliographic Description (`isbd`) stammt, findet man in der nachfolgenden Tabelle.

Pica+-Subfeld	Element einer bekannten Ontologie	Feldinhalt
002@ \$0 Pos. 1 = A,I,K,M und Pos. 2 = a	<code>Dcterms:format</code>	„print“
002@ \$0 Pos. 1 = A,H,I,K,M und Pos. 2 = a	<code>dcterms:medium</code>	„paper“
002@ \$0 Pos. 1 = A,H,I,K,M und Pos. 2 = a	<code>rdf:typ</code> <code>rdf:resource</code>	<code>bibo:book</code>
003@ \$0	<code>dcterms:identifizier</code>	PPN des Titeldatensatzes
003O \$0	<code>dcterms:identifizier</code>	OCLC-Nummer
004A \$0	<code>bibo:isbn10</code>	10-stellige ISBN-Nummer
004A \$A	<code>bibo:isbn13</code>	13-stellige ISBN-Nummer
005A \$0	<code>bibo:issn</code>	ISSN
006G \$0	<code>owl:sameAs</code>	URI des Titels bei der DNB
010@ \$a	<code>dcterms:language</code>	Sprachbezeichnungen
011@ \$a	<code>dcterms:issued</code>	Erscheinungsjahre
019A \$a	<code>dcterms:spatial</code>	Ländercode
021A \$a	<code>dcterms:title</code>	Hauptsachtitel
021A \$a	<code>isbd:P1004</code>	Hauptsachtitel
021A \$d	<code>isbd:P1006</code>	Zusatz
028A \$0	<code>dcterms:creator</code>	Author
028A \$0, 028B \$0, 029A \$0	<code>bibo:authorlist</code>	URIs auf PND
028C \$0	<code>dcterms:contributor</code>	URIs auf PND
032@ \$a	<code>dcterms:edition</code>	Ausgabebezeichnung
032@ \$a	<code>isbd:P1008</code>	Ausgabebezeichnung
033A \$n	<code>dcterms:publisher</code>	Verlag
034D \$a	<code>dcterms:extent</code>	Umfangsangabe
036M \$x	<code>bibo:volume</code>	Zählung Sortierform
041A \$0	<code>dcterms:subject</code>	URI auf GND

Die in den Subfeldern des Pica+-Datensatzes enthaltenen Werte werden bei der Konvertierung zu Objekten der Tripel. In der Mehrzahl sind diese Werte Literale. Doch in einigen Fällen sind diese Werte auch URIs, die beispielsweise auf die DNB verweisen (siehe in der obigen Tabelle 006G \$0, 028A \$0, 028B \$0, 029A \$0, 041A \$0).

Um auch Subfelder in RDF darstellen zu können, für die bisher noch kein geeignetes Vokabular aus einer bereits vorhandenen Ontologie gefunden werden konnte, wurde eine triviale Ontologie „picaplus“ eingeführt. Hiermit werden Prädikate gebildet, die die Kategorie und das Subfeld enthalten. So ergibt sich für das Subfeld \$0 der Kategorie 002@ beispielsweise folgendes Prädikat:

```
<picaplus:tag002at_0>Aau</picaplus:tag002at_0>
```

*Als Beispiel folgt nun die Darstellung der Titeldaten zur ISBN 28068667 in RDF/XML-Format:*

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:picaplus="http://lod.hebis.uni-frankfurt.de/daten/terms/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:bibo="http://purl.org/ontology/bibo/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:wdrs="http://www.w3.org/2007/05/powder-s#"
  xmlns:isbd="http://iflastandards.info/ns/isbd/elements/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  <rdf:Description rdf:about="http://resolver.hebis.de/resource/28068667">
    <rdf:type
rdf:resource="http://purl.org/vocab/frbr/core#Manifestation"/>
    <wdrs:describedby
rdf:resource="http://resolver.hebis.de/catalog/html/28068667"/>
    <picaplus:tag001at_0>3</picaplus:tag001at_0>
    <picaplus:tag001A_0>0030:16-11-11</picaplus:tag001A_0>
    <picaplus:tag001B_0>1999:16-11-11</picaplus:tag001B_0>
    <picaplus:tag001B_t>23:13:59.000</picaplus:tag001B_t>
    <picaplus:tag001C_0>0030:16-11-11</picaplus:tag001C_0>
    <picaplus:tag001C_t>08:50:51.000</picaplus:tag001C_t>
    <picaplus:tag001D_0>0030:16-11-11</picaplus:tag001D_0>
    <picaplus:tag001U_0>utf8</picaplus:tag001U_0>
    <picaplus:tag001X_0>0</picaplus:tag001X_0>
    <picaplus:tag002at_0>Aau</picaplus:tag002at_0>
    <dcterms:medium>paper</dcterms:medium>
    <dcterms:format>print</dcterms:format>
    <rdf:type
rdf:resource="http://purl.org/dc/terms/BibliographicResource"/>
    <rdf:type rdf:resource="http://purl.org/ontology/bibo/Book"/>
    <dcterms:identifrier>280686676</dcterms:identifrier>
    <picaplus:tag0030_a>OCoLC</picaplus:tag0030_a>
    <dcterms:identifrier>760128280</dcterms:identifrier>
    <bibo:isbn13>978-3-8274-1800-5</bibo:isbn13>
    <picaplus:tag007A_a>HEB</picaplus:tag007A_a>
    <picaplus:tag007A_0>280686676</picaplus:tag007A_0>
    <picaplus:tag009P_q>text/html</picaplus:tag009P_q>
    <picaplus:tag009P_u>http://deposit.d-nb.de/cgi-bin/dokserv?
id=2922186&prov=M&dok_var=1&dok_ext=htm</picaplus:tag009P_u>
    <picaplus:tag009P_3>Inhaltstext</picaplus:tag009P_3>
    <picaplus:tag009P_A>2</picaplus:tag009P_A>
    <picaplus:tag009P_m>V:DE-604</picaplus:tag009P_m>
    <picaplus:tag009P_q>application/pdf</picaplus:tag009P_q>
    <picaplus:tag009P_u>http://scans.hebis.de/HEBCGI/show.pl?
28068667_toc.pdf</picaplus:tag009P_u>
    <picaplus:tag009P_3>Inhaltsverzeichnis</picaplus:tag009P_3>
    <picaplus:tag009P_A>2</picaplus:tag009P_A>
    <dcterms:language>ger</dcterms:language>
    <dcterms:issued>2011</dcterms:issued>
```

```

    <picaplus:tag013H_0>z</picaplus:tag013H_0><dcterms:title>Biochemie</d
cterms:title>
    <isbd:P1004>Biochemie</isbd:P1004>
    <picaplus:tag021A_h>Jeremy M. Berg ; John L. Tymoczko ; Lubert
Stryer</picaplus:tag021A_h>
    <picaplus:tag022A_a>Biochemistry</picaplus:tag022A_a>
    <picaplus:tag022A_r>dt.</picaplus:tag022A_r>
    <bibo:authorlist>
      <rdf:Seq>
        <rdf:li rdf:resource="http://d-nb.info/gnd/12460109X"/>
        <rdf:li rdf:resource="http://d-nb.info/gnd/124601103"/>
        <rdf:li rdf:resource="http://d-nb.info/gnd/124601197"/>
      </rdf:Seq>
    </bibo:authorlist>
    <bibo:edition>6. Aufl., korrigierter Nachdr.</bibo:edition>
    <isbd:P1008>6. Aufl., korrigierter Nachdr.</isbd:P1008>
    <picaplus:tag033A_p>Heidelberg</picaplus:tag033A_p>
    <dcterms:publisher>Spektrum, Akad. Verl.</dcterms:publisher>
    <dcterms:extent>XXXVIII, 1224 S.</dcterms:extent>
    <picaplus:tag034M_a>Ill., zahlr. graph. Darst.</picaplus:tag034M_a>
    <picaplus:tag041A_9>085010251</picaplus:tag041A_9>
    <dcterms:subject>http://d-nb.info/gnd/4006777-4</dcterms:subject>
    <picaplus:tag041A_s>Biochemie</picaplus:tag041A_s>
    <picaplus:tag041A_a>f Lehrbuch</picaplus:tag041A_a>
    <picaplus:tag044N_a>Protease</picaplus:tag044N_a>
    <picaplus:tag045E_c>570</picaplus:tag045E_c>
    <picaplus:tag045F_e>DDC22ger</picaplus:tag045F_e>
    <picaplus:tag045F_a>572</picaplus:tag045F_a>
    <picaplus:tag045F_a>572</picaplus:tag045F_a>
    <picaplus:tag045Z_a>VK 8500</picaplus:tag045Z_a>
    <picaplus:tag045Z_a>WD 4010</picaplus:tag045Z_a>
    <picaplus:tag045Z_a>CZ 1100</picaplus:tag045Z_a>
    <picaplus:tag045Z_a>WD 4000</picaplus:tag045Z_a>
    <picaplus:tag046M_a>Literaturangaben</picaplus:tag046M_a>
  </rdf:Description>
</rdf:RDF>

```

### Weitere bisher durchgeführte Maßnahmen zur Veröffentlichung der HeBIS-Titeldaten als Linked-Data

Mit dem oben genannten Programm „rdf“ ist es nun auch möglich einen Gesamtabzug aller HeBIS-Titeldaten im RDF/XML-Format zu erstellen und diesen in den Quad-Store eines Virtuoso-Datenbank-Servers einzuspielen. Ein solcher ist zu Testzwecken bereits auf einem Windows-PC installiert worden. Über den SPARQL-Endpoint dieses Virtuoso-Datenbank-Servers können dann auch komplexere Recherchen mittels SPARQL-Abfragen durchgeführt werden. Auch ein gezieltes Ändern und Auslesen der im Quad-Store enthaltenen Daten ist hiermit möglich.