Empfehlung für die RDF-Repräsentation bibliografischer Daten (Textressourcen)

Bei dieser Webseite handelt es sich um eine inhaltsgleiche Webversion zur PDF-Publikation "Empfehlungen zur RDF-Repräsentation bibliografischer Daten / DINI-AG KIM Gruppe Titeldaten. (DINI-Schriften - 14) Version 2.0. November 2018.", urn:nbn:de:kobv:11-110-18452/2153.3-7

- Versionshistorie
- Über die DINI-AG KIM Gruppe Titeldaten
- Vorgehen und Konventionen
    - Erläuterungen zur Vorgehensweise
    - Nachnutzung von RDF-Elementen
    - Darstellung von RDF-Beispielen
    - Mappings
    - Wenn URI-Referenzen nicht in allen Fällen verfügbar sind
        - Option 1 "Hash-URI"
        - Option 2 "blank node"
    - Lizenzangaben für bibliografische Daten
- Anhang

Versionshistorie

23.11.2018	Version 2.0 (diese Webseite) urn:nbn:de:kobv:11-110-18452/2153.3-7 Überarbeitungen im Zuge der Einführung von Resource Description and Access (RDA) an vielen Institutionen des deutschen Sprachraums, Anpassungen an Erkenntnisse aus der Praxis der beteiligten RDF-Datenproduzenten Änderungen gegenüber Version 1.1 (DIFF-Ansicht)
02.06.2014	Version 1.1 (Webversion) urn:nbn:de:kobv:11-100217673 Korrektur des RDA Namespaces und der Base Domain
30.09.2013	Version 1.0 (Webversion) urn:nbn:de:kobv:11-100212769

- Versionshistorie
- Über die DINI-AG KIM Gruppe Titeldaten
- Vorgehen und Konventionen
    - Erläuterungen zur Vorgehensweise
    - Nachnutzung von RDF-Elementen
    - Darstellung von RDF-Beispielen
    - Mappings
    - Wenn URI-Referenzen nicht in allen Fällen verfügbar sind
        - Option 1 "Hash-URI"
        - Option 2 "blank node"
    - Lizenzangaben für bibliografische Daten
- Anhang

Über die DINI-AG KIM Gruppe Titeldaten

Die Gruppe Titeldaten wurde im Januar 2012 gegründet, seit April 2012 agiert sie als Untergruppe des Kompetenzzentrums Interoperable Metadaten (DINI-AG KIM). KIM ist eine Arbeitsgruppe der Deutschen Initiative für Netzwerkinformation e. V. (DINI e.V.). Vertreten sind alle deutschen Bibliotheksverbünde, die Deutsche Nationalbibliothek, die Österreichische Bibliothekenverbund und Service GmbH, die Schweizerische Nationalbibliothek sowie einige weitere interessierte und engagierte Kolleginnen und Kollegen mit entsprechender Expertise. Die Moderation und Koordination liegt bei der Deutschen Nationalbibliothek.

Die Gruppe Titeldaten dient als Forum für den fachlichen Austausch über die Repräsentation von bibliografischen Daten als Linked Data. Die Teilnehmerinnen und Teilnehmer teilen ihre Erfahrungen, sammeln Best Practices und markieren Fallstricke in Bezug auf die Transformation vorhandener Titeldaten nach RDF.

Ergebnis dieser Diskussionen sind die "Empfehlungen zur RDF-Repräsentation bibliografischer Daten", die nun in der Version 2.0 vorliegen. Die Empfehlungen dokumentieren Lösungen und zeigen mögliche Alternativen auf. Sie sollen zum einen Neueinsteigern auf dem Gebiet die Arbeit erleichtern und zum anderen zu einer Harmonisierung der RDF-Repräsentationen von Titeldaten im deutschsprachigen Raum beitragen. Durch die Etablierung eines “Quasi-Standards” soll die Interoperabilität bibliografischer Linked-Open-Data-Publikationen im deutschsprachigen Raum gewährleistet werden.

Die vorliegende Version 2.0 der Empfehlungen berücksichtigt die Entwicklungen der letzten Jahre. Den initialen Anlass für die Überarbeitung gab die Einführung des Katologisierungsregelwerks *Resource Description and Access* (RDA) in großen Teilen der deutschsprachigen Bibliothekslandschaft in 2015 /2016.

Vorgehen und Konventionen

Erläuterungen zur Vorgehensweise

Die DINI-AG KIM Gruppe Titeldaten analysierte zunächst, welche Metadaten zur Beschreibung einer bibliografischen Ressource erforderlich sind.

Die vorliegenden Empfehlungen beinhalten die Modellierung eines Kernelementsets, das für die eindeutige Identifizierung einer bibliografischen Ressource als notwendig erachtet wird.

Im ersten Schritt wird ein Kernelementset für textuelle Ressourcen beschrieben. Die Beschreibung von Kernelementen für weitere Objekttypen (Film, Musik, etc.) sollen in späteren Versionen folgen.

Je nach Anwendungsfall können Anforderungen bestehen, zusätzliche Informationselemente mit weiteren RDF-Properties auszuweisen oder eine größere Auswahl an Informationselementen in die RDF-Repräsentation eines Titeldatensatzes mit aufzunehmen. Dies liegt im Ermessen der einzelnen Datenproduzenten. Die Erweiterten Empfehlungen für Textressourcen der DINI-AG KIM Gruppe Titeldaten sammeln Best Practices, die über das Kernelementeset hinausgehen.

Nachnutzung von RDF-Elementen

Die vorliegende Modellierung nutzt die unten tabellarisch aufgeführten RDF-Element Sets nach.

Die Auswahl richtet sich nach dem Grad der Verbreitung. Das heißt konkret: Im ersten Schritt wird das Dublin Core-Vokabular nach geeigneten Elementen untersucht, da dieses auch außerhalb der Bibliothekswelt stark verbreitet ist. In den Versionen 1.x dieser Empfehlungen wurden (bibliotheks-)spezifischere Elemente bevorzugt aus der Bibliographic Ontology entnommen. Mit Version 2.0, nach der Einführung der Resource Description and Access (RDA) an vielen Bibliotheken im deutschen Sprachraum, werden weitere Spezifizierungen auch mit den "unconstrained" Properties des RDA Element Sets empfohlen. Deren Verwendung ist dabei unabhängig vom zugrunde liegenden Regelwerk der damit abzubildenden Daten möglich. Der Grund für die Wahl der "unconstrained" Variante der RDA-Element Sets ist der Wunsch, keine Ebenen des IFLA Library Reference Model (IFLA-LRM) zu implizieren, damit diese Empfehlungen möglichst breit anwendbar sind.

Präfix	Element Set
dct	DCMI Metadata Terms
dc	Dublin Core Metadata Element Set, Version 1.1
rdau	RDA Unconstrained properties
bibo	Bibliographic Ontology
marcRole	MARC Code List for Relators
umbel	Upper Mapping and Binding Exchange Layer
schema	Schema.org
isbd	International Standard Bibliographic Description

Darstellung von RDF-Beispielen

Illustrierende Beispiele werden in Terse RDF Triple Language-Serialisierung (Turtle) notiert.

Mappings

- Konzeptuelles Mapping (Inhalt - RDF)
  - Dieses Mapping bleibt auf der konzeptuellen Ebene, das heißt die vorliegenden Empfehlungen sollen unabhängig von konkreten bibliothekarischen Datenformaten anwendbar sein. Deshalb werden die einzelnen Informationselemente natürlichsprachig benannt und nicht anhand von MARC-Feldern o.ä. identifiziert.
- MARC 21-RDF-Mapping

Die Beispielimplementierungen (siehe Anhang) beziehungsweise die aktiven Datenproduzenten können bei Bedarf Auskunft über Mappings aus weiteren bibliothekarischen Datenformaten geben.

Wenn URI-Referenzen nicht in allen Fällen verfügbar sind

Viele Datenproduzenten haben mit dem Problem zu tun, dass einzelne Informationselemente sowohl als URI-Referenz als auch als Literal in den Daten vorliegen können.

Beispiel

Das Element "Schöpfer" wird nach aktueller Erfassungspraxis der Beispielinstitution über einen Link auf die Gemeinsame Normdatei (GND) identifiziert, wurde aber in einer bestimmten Generation von Altdaten als Literal erfasst.

Da es sich um dasselbe Informationselement handelt, sollte für beide Fälle dieselbe rdf:Property verwendet werden. Andernfalls ist für den Nutzer der Daten Hintergrundwissen über der Modellierung der Daten notwendig, um beide Formen gleichberechtigt auswerten zu können.

Es ist außerdem nicht ratsam, eine Property im selben Datenset sowohl mit URIs in Objektposition als auch mit Literalen in Objektposition zu verwenden. Das gilt auch für Properties, deren Definition keine Aussage zum erwarteten Objekttyp macht (wenn zum Beispiel keine rdfs:range-Angabe vorliegt). Deshalb wird hier empfohlen, stattdessen durchgängig URIs beziehungsweise unter Umständen auch Blank Nodes in Objektposition zu verwenden.

Mit anderen Worten sollte bei bloßen Literalangaben ohne IDs in den Quelldaten die Entität, für die das Literal steht, in RDF modelliert und referenzierbar gemacht werden. Die Modellierung sollte mindestens eine Property rdfs:label (oder vergleichbar) enthalten. Für die Referenzierung werden hier zwei Optionen skizziert:

1. Generierung eines Hash-URIs für die Entität
2. Arbeiten mit einem blank node

Option 1 "Hash-URI"

Ein Fragmentsbezeichner, eingeleitet durch das Hash-Zeichen ("#"), kann einem URI angefügt werden. Die Kombination bildet dann einen weiteren URI, der etwas eindeutig identifizieren kann. Über das was, er identifizieren soll, kann dann ebenfalls unter dem Ausgangs-URI eine Beschreibung geliefert werden.

Beispiel Option 1 "Hash-URI"

```
@prefix dct: <http://purl.org/dc/terms/> .
@prefix rdfs: http://www.w3.org/2000/01/rdf-schema#> .

<http://id.meineInstitution.de/12345> dct:creator <http://id.meinebibliothek.de/12345#creator_Bergmann_Bertram>
.
<http://id.meineInstitution.de/12345#creator_Bergmann_Bertram> rdfs:label "Bertram Bergmann" .
```

Bei der Erzeugung der RDF-Daten sollte der Fragmentbezeichner so generiert werden, dass er sich bei ggf. anfallender erneuter Erzeugung der Daten (falls diese regelmäßig aus einem anderen Format konvertiert werden) möglichst nicht ändert. Diese Lösung ist der Option "blank node" vorzuziehen, da so die Entität hinter dem Literal nicht nur temporär und im Kontext der Datenversion referenzierbar ist (wie beim blank node, siehe dort), sondern längerfristig. Das kann auch Mappingprojekte begünstigen, die auf eine Ergänzung der fehlenden Links hinarbeiten.

Option 2 "blank node"

Ein "leerer Knoten" beziehungsweise englisch "blank node" wird im RDF-Modell eine nicht-global-referenzierbare Ressource genannt.

Beispiel Option 2 "blank node"

```
@prefix dct: <http://purl.org/dc/terms/> .
@prefix rdfs: http://www.w3.org/2000/01/rdf-schema#> .

<http://id.meineInstitution.de/12345> dct:creator [
    rdfs:label "Bertram Bergmann" .
]
```

Die Verwendung von blank nodes mindert die Nutzbarkeit der Daten (gegenüber URIs), da die betroffene Entität nicht eindeutig referenziert und somit nicht verlinkt werden kann. Innerhalb einer Datenversion (zum Beispiel *Datendumpdatei vom 16.05.2017*) wird die Referenzierbarkeit der im blank node beschriebenen Entität zwar hergestellt über einen nur für diese Datenversion gültigen temporären Identifier (siehe Beispiel unten, das semantisch dem Beispiel oben entspricht). Allerdings handelt es sich bei einem solchen Identifier weder um eine stabile ID noch um einen HTTP-URI. Somit wird damit keine Referenzierbarkeit durch externe Datensets hergestellt.

Beispiel Option 2 "blank node", Darstellung mit temporärer ID

```
@prefix dct: <http://purl.org/dc/terms/> .
@prefix rdfs: http://www.w3.org/2000/01/rdf-schema#> .

<http://id.meineInstitution.de/12345> dct:creator _:node1b6mbf4tqx36921502 .
_:node1b6mbf4tqx36921502 rdfs:label "Bertram Bergmann" .
```

Lizenzangaben für bibliografische Daten

Betreffend die Lizensierung bibliografischer Daten sollte sich an der Empfehlungen zur Öffnung bibliothekarischer Daten der DINI-AG KIM Gruppe Lizenzen orientiert werden. Die vergebene Lizenz sollte in RDF transportiert werden. Es wird empfohlen, dies im Rahmen von Datenset-Beschreibungen zu tun. In der W3C Empfehlung "Data on the Web Best Practices" sind im Abschnitt 8.2. Metadata entsprechende Empfehlungen enthalten. Die Zuordnung der Datenset-Zugehörigkeit eines einzelnen RDF-Dokuments setzt die Beachtung der URI-Design-Empfehlungen, das heißt die klare Unterscheidung der URIs für ein Ding und seine Beschreibung, voraus. Mit Hilfe des URIs der Beschreibung und der Property void:inDataset kann das RDF-Dokument als Teil eines Datensets ausgezeichnet werden.

Beispiel Datenset-Zugehörigkeit und Lizenz des Datensets

```
@prefix void: <http://rdfs.org/ns/void#> .
@prefix ex:    <http://www.example.org/> .
@prefix dct: <http://purl.org/dc/terms/> .


<http://d-nb.info/1118539508>
    void:inDataset <http://d-nb.info/dnb-all#dataset> .

<http://d-nb.info/dnb-all#dataset> a ex:Dataset ;
    dct:license <http://creativecommons.org/publicdomain/zero/1.0/> .
```

Anhang

- Mitwirkende der DINI-AG-KIM Gruppe Titeldaten:
  - Barbara Block - GBV
  - Iris Hausmann - BSZ
  - Sarah Hartmann - DNB
  - Jana Hentschke - DNB
  - Reinhold Heuvelmann - DNB
  - Andreas Kahl - BSB
  - Cornelia Katz - BSZ
  - Carsten Klee - ZDB
  - Adrian Pohl - hbz
  - Stefanie Rühle - SUB Göttingen
  - Christiane Schmidt - Schweizerische Nationalbibliothek
  - Thomas Striffler - HeBIS
  - Lars G. Svensson - DNB
- Mitwirkende an früheren Versionen dieser Empfehlungen
  - Stefan Brecheisen - BVB
  - Pascal Christoph - hbz
  - Julia Hauser - DNB
  - Dieter Janka - BSZ
  - Verena Schaffner - OBVSG
  - Stephani Scholz - hbz

- URI-Design-Empfehlungen

- Best Practices / Beispielimplementierungen:
  - Linked Data Service der Deutschen Nationalbibliothek
  - Linked Data Service der Zeitschriftendatenbank
  - lobid-resources – der hbz-Verbundkatalog als LOD, hbz — Hochschulbibliothekszentrum des Landes NRW
  - LinkedOpenData-Service des B3Kat
  - HeBIS Linked Open Data Services
  - SWB-Verbunddatenbank als Linked Open Data