

HANDREICHUNG LINKED OPEN DATA

Was macht Open Data zu Linked Open Data?

Nach dem 5-Sterne-Modell von Tim Berners-Lee ist Linked Open Data der höchste Reifegrad, den offene Daten erlangen können. Denn Linked Open Data erfüllt eine Reihe von Gestaltungsprinzipien für

das Erstellen von strukturierten und maschinenlesbaren Daten. Diese erlauben es, eine Verlinkung zu anderen offenen Daten herzustellen und damit die Daten durch Kontext anzureichern:

1. Die Verwendung von URIs als Bezeichner für Gegenstände

Ein Uniform Resource Identifier (URI) ist eine eindeutige Zeichenfolge, die als Bezeichner/Identifikator für alles verwendet werden kann – unabhängig davon, ob es sich um virtuelle Inhalte, Objekte aus der realen Welt oder abstrakte Konzepte han-

delt. Die Grundvoraussetzung bei der Verwendung von URIs ist, dass die URIs global und persistent sind. Es muss also dauerhaft gewährleistet sein, dass jeder Bezeichner für eine Sache gültig und über die eigene Anwendung hinaus eindeutig ist.

2. Die Verwendung von HTTP URIs, damit die Bezeichner aufgerufen werden können

Das Hypertext Transfer Protocol (HTTP) ist ein Internetprotokoll, das der Übertragung von Daten über ein Rechnernetzwerk dient. Immer wenn eine Webseite in einem Browser aufgerufen wird, wird eine HTTP-Anfrage an den entsprechenden Webserver gesendet, der die angeforderte HTML-Seite

zurück liefert. Wird nun für eine Sache eine URI vergeben, die als URL auflösbar ist, kann die Sache nicht nur eindeutig identifiziert, sondern auch direkt auf sie zugegriffen werden. So können weitere Informationen zu der Ressource hinterlegt und diese damit kontextualisiert werden.

3. Das Hinterlegen nützlicher Informationen unter der Verwendung der Standards RDF und SPARQL

Linked Data verknüpft Daten, die in einer maschinenlesbaren Form vorliegen, mit weiteren Daten, um sie in Kontexte einzu-

betten. Dazu ist die Verwendung eines standardisierten Formats notwendig.

Im Kontext von Linked Open Data wird Resource Description Framework (RDF) genutzt, um Linked Data darzustellen. Mittels RDF können logische Aussagen über **Gegenstände** aus der realen Welt in Form von sogenannten **Tripel** getroffen werden, die sich aus Subjekt, Prädikat und Objekt zusammensetzen. Dabei werden Subjekt und Prädikat stets als URI ausgedrückt, während es sich bei dem Objekt entweder um eine URI oder ein Literal (z.B. Zahlen, Daten oder Zeichenfolgen) handeln kann. Wird das Objekt als URI ausgedrückt, kann es wiederum zum Subjekt eines weiteren Tripels werden. Nimmt man alle Tripel einer RDF-Struktur

zusammen, erhält man einen Graphen. Das folgende Beispiel drückt aus, dass der Titel der Ressource <https://open.rlp.de> **open.rlp** lautet, wobei es sich bei der Zeichenkette **open.rlp** nicht um eine URI, sondern um ein Literal handelt. Das Prädikat ist dabei eine URI, die ein Bestandteil des Dublin Core Vokabulars ist. Dublin Core ist ein Metadatenschema, das der Beschreibung elektronischer Ressourcen dient. Die Prädikats-URI ist als URL auflösbar und weist auf eine Beschreibung des **title** in natürlicher Sprache und ermöglicht es einem Computer, die Angabe als Titel zu verstehen.

<https://open.rlp.de> <http://purl.org/dc/terms/title> „open.rlp“ .

Bei dem Beispiel handelt es sich um ein N-Triple. N-Triple sind ein zeilenbasiertes Reintextformat zur Serialisierung von RDF-Triples.

Die standardisierte Abfragesprache für RDF ist SPARQL. Wegen des beschränkten Umfangs dieser Handreichung wird SPARQL im weiteren Verlauf nicht weiter betrachtet werden.

4. Das Verlinken von anderen URIs in den eigenen Daten

Durch die Nutzung des Datenmodells RDF können Daten mit anderen Daten verknüpft werden. Die Vision ist ein globales Netz-

werk miteinander verknüpfter maschinenlesbarer Informationen, die von allen frei genutzt werden können.

Resource Description Framework (RDF)

RDF bildet einen Grundstein für die Umsetzung von Linked Open Data. RDF als Methode zur Beschreibung von Ressourcen bringt einige Vorteile mit sich:

Geläufige Tabellenkalkulationsprogramme können zur Generierung von Dateien im CSV-Format genutzt werden. Dabei gibt es einige Punkte zu beachten:

- Die Graphstruktur sorgt für Flexibilität, da sie auch nachträglich verändert und angepasst werden kann.
- Das Zusammenführen von Daten-

beständen wird ebenfalls durch die Graphstruktur vereinfacht, da sie nicht zwangsläufig eine festgelegte Struktur mitbringt.

- Durch die Verwendung von globalen Identifikatoren werden Mehrdeutigkeiten vermieden, selbst wenn Datenbestände aus mehreren Quellen zusammengeführt werden.

RDF ist unabhängig von seiner textuellen Repräsentation. Im vorangegangenen Beispiel wurde bereits N-Triple als Form der Serialisierung vorgestellt. Darüber hinaus existieren weitere Formen der Repräsentation:

Turtle: Bei [Turtle](#) handelt es sich um eine vereinfachte Form von N3. Das Format ist leichter zu lesen und zu schreiben als andere Formate, jedoch ist es weniger verbreitet als beispielsweise RDF/XML.

```
@prefix dc: <http://purl.org/dc/terms/> .  
<https://open.rlp.de> dc:title „open.rlp“ .
```

RDF/XML: Bei [RDF/XML](#) handelt es sich historisch betrachtet um das erste Serialisierungsformat von RDF. Wie der Name bereits verrät, werden die Daten in Form eines XML-Dokuments dargestellt.

```
<?xml version="1.0" encoding="utf-8" ?>  
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"   
      xmlns:dc="http://purl.org/dc/terms/">  
  <rdf:Description rdf:about="https://open.rlp.de">  
    <dc:title>open.rlp</dc:title>  
  </rdf:Description>  
</rdf:RDF>
```

/open

JSON-LD: Bei JSON-LD handelt es sich um ein leichtgewichtiges Format für Linked Data. Es ist darauf optimiert, von Menschen gelesen und geschrieben zu werden und basiert auf dem weit verbreiteten JSON-Format.

```
[
  {
    „@id“:“https://open.rlp.de“,
    „http://purl.org/dc/terms/title“:
      [
        {
          „@value“:“open.rlp“
        }
      ]
  }
]
```

RDFa: Darüber hinaus bietet RDFa eine Möglichkeit, Tripel in HTML-Dokumente einzubetten. Beispielsweise basiert unter anderem die Idee von Schema.org darauf, RDFa zu nutzen, um im Internet standardisierte strukturierte Daten zur Verfügung zu stellen. Schema.org bietet eine Vielzahl von Schemata, die genutzt werden können, um Daten für verschiedene Themengebiete zu modellieren.

RDF automatisiert generieren

Das automatisierte Erstellen von RDF ist einem händischen Prozess in jedem Fall vorzuziehen. Zu diesem Zweck steht für die Programmiersprache Python die Bibliothek RDFLib bereit. Sie ermöglicht das Einlesen, Serialisieren und Generieren von RDF. Mit Hilfe der zur Verfügung gestellten Funktionen lassen sich eigene Datenbestände in RDF überführen.

SHACL

Die Shapes Constraint Language (SHACL) dient der Beschreibung von RDF-Graphen. Dabei können Beschränkungen für den Inhalt und die Struktur des Graphen definiert und der Wertebereich eingeschränkt werden. Ein Anwendungsbeispiel dafür bildet der DCAT-AP.de Validator, der Dateien oder RDF-Endpunkte gegen die Spezifikation von DCAT-AP.de prüft.



RheinlandPfalz

MINISTERIUM FÜR ARBEIT,
SOZIALES, TRANSFORMATION
UND DIGITALISIERUNG

Herausgeber

Open-Data-Kompetenzzentrum Rheinland-Pfalz im
Ministerium für Arbeit, Soziales, Transformation
und Digitalisierung des Landes Rheinland-Pfalz

Bauhofstr. 9
55116 Mainz

E-Mail: cc-od@open.rlp.de