

# DCAT-AP.de URI-Konzept

Vorschlag zu technischen, semantischen und organisatorischen Konventionen für URIs auf „DCAT-AP.de“ und auf „GovData“

Version: 1.0

## Metadaten des Dokumentes

<b>Eigenschaft</b>	<b>Wert</b>
Dokumentdatum:	21.03.2019
Status:	veröffentlicht
Version:	1.0
Lizenz:	Creative Commons Namensnennung 4.0 „init[ AG und SID Sachsen für die GKSt GovData“
Ersteller:	Sebastian Sklarß [INIT], Dietmar Gattwinkel [Sachsen], Antje Göldner [GKSt GovData], Christian Horn [GKSt GovData] unter Berücksichtigung zahlreicher Anregungen von Teilnehmern mehrerer Online- und Präsenzabstimmungsrunden
Zugriffs-URI:	<a href="http://dcat-ap.de/def/uriConcept/1.0">http://dcat-ap.de/def/uriConcept/1.0</a>

## Abbildungsverzeichnis

Abbildung 1: Logo des EU Data Catalogue Vocabulary Application Profile.....	4
Abbildung 2: aktuelle Datensatz-URL von GovData.de .....	6
Abbildung 3: Zugriff auf Class- und Property-URI des HTML bzw. RDF .....	11
Abbildung 4: Rolle des URI-Konzeptes im DCAT-AP.de Standard.....	12
Abbildung 5: Logo von GovData.....	13
Abbildung 6: 5 Star Schema .....	13
Abbildung 7: URIs in DCAT-AP .....	15

## Inhaltsverzeichnis

<b>1</b>	<b>Konventionen und Empfehlungen für URIs</b> .....	<b>5</b>
1.1	Allgemeine Konventionen .....	5
1.2	Vorgaben für den Webserver .....	11
<b>2</b>	<b>Weiterführende Details</b> .....	<b>12</b>
2.1	Wozu ein URI-Konzept? .....	12
2.2	DCAT-AP.de vs. GovData.de .....	13
2.3	Linked Data-Prinzipien .....	13
2.4	Arten von URIs .....	14
2.5	URIs als Bestandteil von DCAT-AP .....	15
2.6	URIs als Bestandteil von DCAT-AP.de .....	16
2.7	URIs beim GovData Datenbereitsteller .....	17
<b>3</b>	<b>Übersicht über alle Regeln</b> .....	<b>18</b>
<b>4</b>	<b>Quellen</b> .....	<b>20</b>

## Management Summary

Der IT-Planungsrat hat am 28. Juni 2018 beschlossen, dass dem Datenaustausch zwischen dem Datenportal GovData und anderen Datenportalen eine deutsche Ableitung des europäischen Metadatenstandards DCAT-AP zugrunde gelegt werden soll.

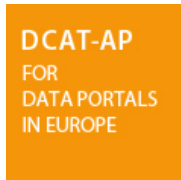


Abbildung 1: Logo des EU Data Catalogue Vocabulary Application Profile

Das internationale Format DCAT-AP wurde zu einem deutschlandweit auf kommunaler, landes- und Bundesebene abgestimmten Format DCAT-AP.de ausgebaut, das das zuvor verwendete Metadatenschema OGD 1.1 abgelöst hat.

Mit diesem URI-Konzept wird eine einheitliche, zukunftsichere Vorgehensweise ermöglicht, um im Internet ergänzende Ressourcen (insbesondere kontrollierte Vokabulare) für die Beschreibung von offenen Verwaltungsdaten auf GovData bereitzustellen.

Dieses Konzept

- regelt die **Nutzung von Uniform Resource Identifiers** im Datenaustausch mit GovData.de unter Nutzung von DCAT-AP.de sowie allgemeine Hinweise und Empfehlungen auch für das eigene Datenportal,
- ermöglicht **sinnvoll benannte, langlebige und verlässliche** URIs, die sich im Laufe der nächsten Jahre möglichst wenig ändern sollen,
- klärt die Namensvergabe zwischen **GovData.de und DCAT-AP.de** und trägt zur einheitlichen Ausgestaltung von URIs innerhalb des gemeinsamen Metadaten-Portalverbundes GovData bei und
- wird durch die Geschäfts- und Koordinierungsstelle GovData bei Bedarf fortgeschrieben.

# 1 Konventionen und Empfehlungen für URIs

## 1.1 Allgemeine Konventionen

Unter Berücksichtigung der beispielhaften Möglichkeiten eines GovData-Datenbereitstellers und des GovData-Datenportals wurden zur Wahrung von Interoperabilität und zum Erreichen von langlebigen URIs die nachfolgenden Konventionen und Empfehlungen festgelegt. Für Datenbereitsteller sind zudem Hinweise für den optimalen Entwurf von URIs für lokale Datensätze enthalten.

Regel 1: Nach dem 01.07.2017 neu angelegte URIs für offene Daten der öffentlichen Verwaltung in Deutschland SOLLEN eine Gültigkeitsdauer von mindestens 10 Jahre haben.

Regel 2: HTTP-Links als URIs MÜSSEN nach Linked Data-Prinzipien verwendet werden, damit URIs aufgelöst werden können. Das bedeutet, dass im Internet zugängliche weitere Metadaten und Informationen (üblicherweise RDF und HTML) hinter den Adresslinks der Konzepte stehen müssen.

Das vorliegende URI-Konzept regelt den technologieneutralen Aufbau einer persistenten GovData-URI und einer persistenten DCAT-AP.de-URI.

DCAT-AP.de wird dabei der Namensraum für alle DCAT-AP.de Codelisten, die speziell für den DCAT-AP.de-basierten Austausch oder für spezielle deutsche Anforderungen in Form von semantischen Ontologien (SKOS Dateien) oder URIs zur Verfügung stehen.

Regel 3: Für URIs semantischer Konzepte unter DCAT-AP.de MUSS folgende Aufbauvorschrift angewendet werden:  
<http://dcat-ap.de/def/{konzept}>; konzept = Codeliste oder URI auf Datei in SKOS:notation.

GovData.de ist der Namensraum für Datensätze und Dokumente sowie für den Bulk-Download aller Datensätze.

Folgende URI-Typen werden für das GovData Portal und für Portale der Datenbereitsteller als **empfohlen** eingeführt:

- dataset (1) – ein Datensatz z.B. auf GovData.de
- def (2) – eine (Code)liste, eine Ontologie oder anderes erweitertes semantisches Konzept steht hinter der URI zur Verfügung
- bulk (2) – ein Bulkdownload steht unter der URI zur Verfügung

Regel 4: Für URIs unter GovData und Portalen anderer Datenbereitsteller SOLL die folgende Ausbauvorschrift angewendet werden:  
<http://portal.de/{type}/{child-id}>; {type=def, dataset, bulk}.

Ein Datenbereitstellerportal soll also z.B. eigene semantische Konzepte ebenfalls unterhalb von /def/ ablegen und etwaige Bulk-Dateien mit „bulk“ folgendermaßen kennzeichnen:  
<http://datenbereitsteller.de/bulk/20170413.rdf>.

Regel 5: Alle verfügbaren Repräsentationen SOLLEN durch dedizierte Erweiterungen der URI direkt zugänglich sein.

GovData sowie andere Datenbereitsteller und Betreiber von Datenportalen bieten aktuell HTML-Sichten, in denen alle beim Datenbereitsteller verfügbaren Repräsentationen als Links aufgeführt sind.

Regel 6: In Fällen, in denen mehr als eine Repräsentations-URI angeboten wird, SOLL es auch möglich sein, mit Content Negotiation auf beide zuzugreifen.

Weitere Ausführungen zu Best Practices der Webservereinstellungen für die Auslieferung von RDF-Vokabularen sind im Kapitel 1.2 „Vorgaben für den Webserver“ aufgeführt.

Regel 7: Bestandteile von URIs SOLLEN keine kurzlebigen Informationen enthalten, die ihre Auflösbarkeit zeitlich begrenzen.

GovData sowie andere Datenbereitsteller und Betreiber von Datenportalen sollen z. B. keine Session-IDs in den URIs ihrer Systeme verwenden, wenn auf diese URIs von anderen Systemen aus dauerhaft stabil verwiesen werden soll.

Regel 8: URIs für offene Datensätze SOLLEN technologieneutral formuliert sein und keine Bestandteile enthalten, die z.B. auf das Veröffentlichungssystem verweisen.

URIs wie „SharedDocs“ oder Wiederholungen der Dateiendung im Pfad sowie andere technische Begriffe sollten vermieden werden.



Abbildung 2: aktuelle Datensatz-URL von GovData.de

Regel 9: URIs SOLLEN verständlich und menschenlesbar sein.

Anders als Web-URLs müssen GovData-URIs selten von Hand eingegeben oder per Telefon weitergegeben werden. Dennoch sollten sie kurz und prägnant unter Wahrung der erlaubten Zeichen des Zeichensatzes beschreiben, was sie identifizieren/auflisten/definieren.

Regel 10: Um DCAT-AP-Eigenschaften mit URIs zu benennen, SOLL auf Englisch zurückgegriffen werden, sofern es sich um einen sinnvoll ins Englische übersetzbaren Begriff handelt.

Zum Beispiel würde der Begriff „Leistungskatalog LeiKa“ (LeiKa) auf Deutsch als URI verwendet werden (`dct:conformsTo/leika/`), während für Konzepte wie „Bundesland“ (state) oder „Häufigkeit“ (frequency) gängige englischsprachige Übersetzungen des deutschen Begriffs existieren.

Regel 11: Abgesehen von Personen der Zeitgeschichte MÜSSEN URIs anonymisiert sein und dürfen keine Personennamen enthalten.

Regel 11 adressiert Anforderungen an den Datenschutz.

Regel 12: Bereits als ID bestehende andere persistente URIs SOLLEN nachgenutzt werden, wenn dem keine schwerwiegenden Gründe entgegenstehen.

Selbst wenn dadurch andere Regeln verletzt werden, steht die Nachnutzbarkeit und Nachverfolgbarkeit zum ID-gebenden System VOR der Wohlgeformtheit der URI. DCAT-AP.de definiert ausnahmsweise eigene URI-Räume für Lizenzen, da die Klarheit bei rechtlichen Fragestellungen für die erfolgreiche Nachnutzung von offenen Daten elementar ist.

Regel 13: Es SOLL bei Wertelisten mit Releasezyklen außerhalb des DCAT-AP Standards für jede Version ein Versions-URI bereitgestellt werden, unter der diese erreichbar ist.

Die gute Praxis der Bereitstellung zweier verschiedener URIs für dynamische Wertelisten SOLL berücksichtigt werden: Zum Beispiel kann zusätzlich zur letzten gültigen Lizenzliste (<http://dcat-ap.de/def/licenses>) auch auf die versionierte Liste unter <http://dcat-ap.de/def/licenses/1.2/> zugegriffen werden.

Regel 14: Es SOLL eine nicht-versionierte URI bereitgestellt werden, welcher jeweils auf die aktuell gültige Version einer Werteliste verweist. Beispiel: <http://dcat-ap.de/def/licenses>.

Damit kann die Liste der Lizenzen zukünftig losgelöst vom DCAT-AP.de-Standard in einfacher Weise um heute noch unbekanntes Lizenzen erweitert werden. Im gesamten GovData-Verbund gilt zukünftig die im Konventionenhandbuch festgelegte Version der Lizenzliste. Mittels versionierter URI kann darauf zugegriffen werden.

Des Weiteren sind verschiedene Versionen von GovData-Lizenzcodes unter Verwendung der URI zu versionieren.

Beispiel: <http://dcat-ap.de/def/licenses/1.0/cc-by-de/3.0> und <http://dcat-ap.de/def/licenses/1.2/cc-by-de/3.0>.

Regel 15: Der Singular SOLL verwendet werden, um ein bestimmtes physisches oder abstraktes „echtes Ding“ zu bezeichnen, also z.B. „/dataset/3434“.

Analog zu Best Practices der UML-Modellierung sollten Objekte aus der „echten Welt“ sowie UML-Klassennamen und ihre technischen Implementierungen und andere URI-Bestandteile wenn möglich im Singular formuliert sein.

Regel 16: Der Plural SOLL angewendet werden, um eine Liste von Dingen zu benennen, also z.B. <http://dcat-ap.de/def/hashAlgorithms/>.

Codelisten und anderen „Listen von echten Dingen“ kann man so leichter ansehen, dass sie eine Ansammlung von mehreren gleichwertigen Begriffen darstellen.

Regel 17: Groß-/Kleinschreibung SOLL nicht als Unterscheidungsmerkmal zwischen URIs verwendet werden.

Die Groß- und Kleinschreibung soll nicht als Unterscheidungsmerkmal für vermeintlich verschiedene URIs gesetzt werden, da die Case-Sensitivität betriebssystemspezifisch unterschiedlich umgesetzt ist. Zur besseren Lesbarkeit wird dennoch die Verwendung von lowerCamelCase empfohlen.

Regel 18: URIs SOLLEN außerhalb des Domain-Namens in lowerCamelCase geschrieben sein.

Zur besseren Verständlichkeit in deutscher Sprache SOLL lowerCamelCase verwendet werden.

Regel 19: URIs zu Linked Data Datensätzen SOLLEN auf die Darstellung von Zeichen als Prozentdarstellung verzichten, sie SOLLEN also nur aus Ziffern, Buchstaben und den Sonderzeichen „-“, „.“, „\_“ und „~“ bestehen. Die reservierten Zeichen SOLLEN entsprechend ihrer Bedeutung im Dokumentenpfad eingesetzt werden. (Siehe: <https://de.wikipedia.org/wiki/URL-Encoding>)

Zur besseren Kompatibilität und zur Reduktion von Komplexität soll auf den üblich eingeschränkten Zeichensatz für englische Dateinamen zurückgegriffen werden. IRIs<sup>1</sup> SOLLEN daher nicht verwendet werden.

---

<sup>1</sup> <https://tools.ietf.org/html/rfc3987>, sowie diskutiert in <https://www.w3.org/International/O-URL-and-ident.html>



## GovData-Lizenzcodes in URIs

Obenstehende Regeln machen Vorgaben für die URIs der Liste der Lizenzen:

1. Versionen von Lizenzen in „LizenzURI“

Beispiel: Die Datenlizenz Deutschland Namensnennung 2.0 bekommt folgende Lizenzcode: **dl-by-de/2.0**

Die Version wird dabei in der URI geführt. Sofern eine Erläuterung auf Drittwebseiten nicht verfügbar ist, wird für HTML-Browser per Content Negotiation der Inhalt der aktuellen GovData-Webseite angezeigt.

2. Lizenzcode als „UniqueID“ der Lizenz

Der Lizenzcode, der bereits wesentliche Merkmale der Lizenz erkennen lässt, wird in Form einer URI gefasst, die den URI-Bildungsregeln (lowerCamelCase, Versionierung über URI) entspricht.

Zusätzlich werden weitere Konventionen, angelehnt an den von Creative Commons bekannten Lizenzbeschreibungsbaukasten, eingeführt:

Regel 20: Lizenzen mit obligatorischer Namensnennung des Lizenzgebers **SOLLEN** den Bestandteil „-by“ im Lizenzcode führen.

Beispiel: <http://dcat-ap.de/def/licenses/cc-by-de>

Regel 21: Lizenzen mit Zero-Kennung **SOLLEN** den Bestandteil „-zero“ im Lizenzcode führen.

Beispiel: <http://dcat-ap.de/def/licenses/cc-zero>

Regel 22: Lizenzen, die eine Bearbeitung der Daten untersagen, **SOLLEN** den Bestandteil „-nd“ im Lizenzcode führen.

Beispiel: <http://dcat-ap.de/def/licenses/cc-nd-de>

Regel 23: Lizenzen mit Share-Alike Attribution **SOLLEN** den Bestandteil „-sa“ im Lizenzcode führen.

Beispiel: <http://dcat-ap.de/def/licenses/cc-sa-de>

Regel 24: Lizenzen mit „Nicht für kommerzielle Nutzung“ Attribution **SOLLEN** den Bestandteil „-nc“ im Lizenzcode führen.

Beispiel: <http://dcat-ap.de/def/licenses/dl-by-nc-de/1.0>

Zusätzlich erfolgt eine logische und syntaktische Trennung von Version und Lizenznamen unter Berücksichtigung der Regel 14.

Regel 25: Versionen von Lizenzen **SOLLEN** über die URI abgebildet werden.

Beispiel: <http://dcat-ap.de/def/licenses/apache/2.0>

Portierungen einer Lizenz in nationale Rechtsräume, im GovData-Fall überwiegend den deutschen Rechtsraum „de“, werden im Lizenzcode verschlüsselt.

Regel 26: Die Länderkennung des Rechtsraumes einer Lizenz (die Portierung in eine Sprache) SOLL über die URI abgebildet werden.

Beispiel für eine portierte Lizenz: <http://dcat-ap.de/def/licenses/cc-by-de/3.0>

Regel 27: Die Reihenfolge der generellen Aufbauvorschrift von Lizenz-URIs SOLL wie folgt umgesetzt werden: Lizenzakronym-zero-ByClause-NonCommercial-ShareAlike-NoDerivativeWork-Portierungsländercode/VersionMajor.VersionMinor.

Beispiel: <http://dcat-ap.de/def/licenses/cc-by-nc-sa-de/3.0>

Beispiel: <http://dcat-ap.de/def/licenses/cc-by-nc-nd-de/3.0>

## 1.2 Vorgaben für den Webserver

Anfragende Systeme können HTML-Browser, aber auch RDF-Crawler sein.

Es ist im Bereich des semantischen Webs üblich, folgendes Verhalten eines Servers in Bezug auf ein anfragendes System zu fordern: Content Negotiation (Regel 6), Redirects, Rewrites und das Ausliefern der jeweils aktuellsten Version unter einer stabilen URI (Regel 14). Um eine URI gleichzeitig per Content Negotiation auf zwei unterschiedliche auszuliefernde Inhalte zeigen zu lassen, muss vom anfragenden System die „Agent“-Kennung und das http-ACCEPT-Attribut ausgewertet sowie mit Rewrite ein entsprechender Redirect veranlasst werden. Im Ergebnis wird je nach Kennung (application/rdf+xml oder text/html) eine HTML-Datei oder eine RDF-Datei unter derselben Konzept-URI ausgeliefert, ohne dass dafür die Angabe einer Dateiendung oder eines Dateinamens notwendig ist.

Das Verfahren und dafür notwendige Einträge in der .htaccess-Datei sind in den Varianten Recipe 1-6 in der W3C Working Group Note „Best Practice Recipes for Publishing RDF Vocabularies<sup>2</sup>“ detailliert beschrieben. Zum Beispiel zeigt Abbildung 3 den Zugriff auf einzelne Klassen und Properties eines HTML-Dokuments bzw. RDFs über Class- und Property-URIs.

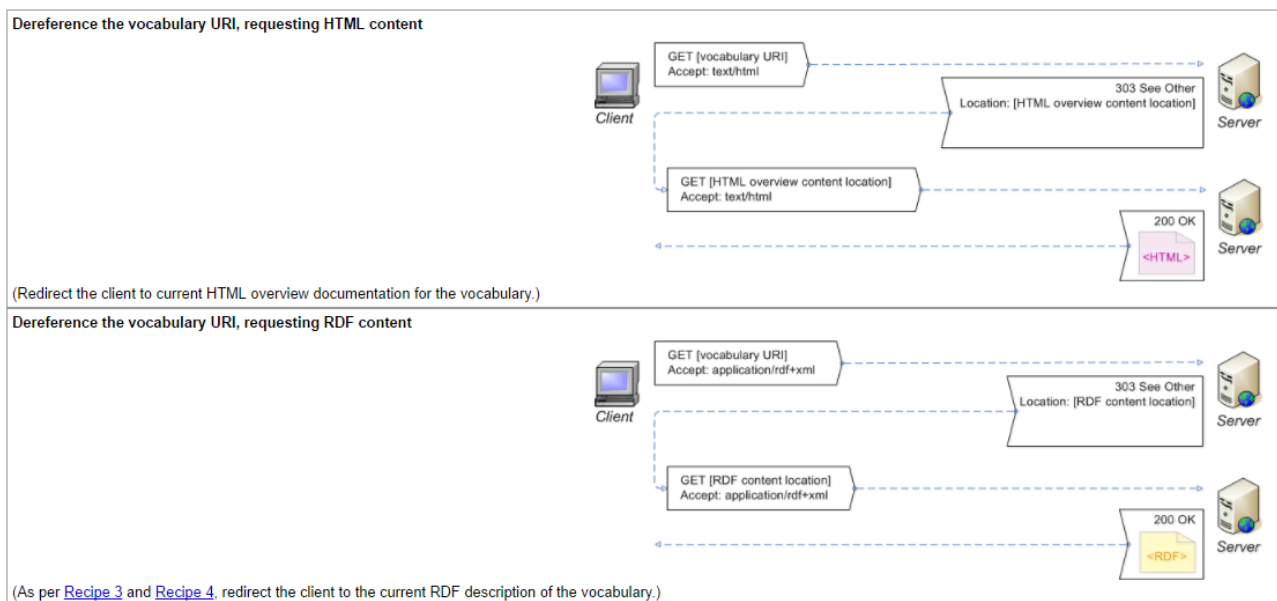


Abbildung 3: Zugriff auf Class- und Property-URI des HTML bzw. RDF

<sup>2</sup> W3C Working Group Note, 28.08.2008, Best Practice Recipes for Publishing RDF Vocabularies, <https://www.w3.org/TR/swbp-vocab-pub/>

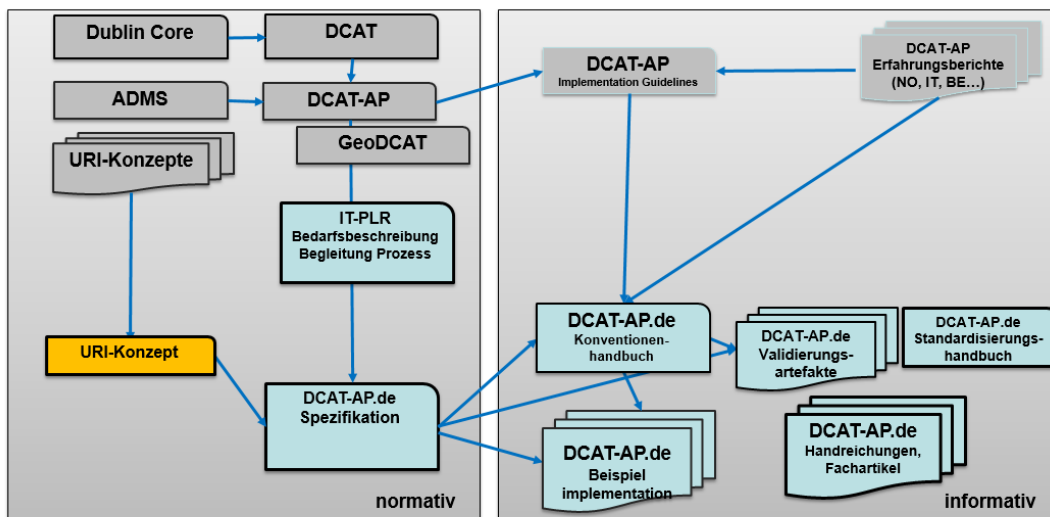
## 2 Weiterführende Details

### 2.1 Wozu ein URI-Konzept?

Um möglichst wenig Änderungsbedarf an der Quelle und im GovData-Portal hervorzurufen, folgen die URIs international erfolgreich erprobten Regeln und Entwurfsmustern.

Dazu wurden, wie in Abbildung 4 dargestellt, URI-Konzepte anderer Länder (UK, Australien, Niederlande) nach Konventionen und Empfehlungen untersucht, Best Practices identifiziert, auf Deutsch übersetzt und für GovData zusammen mit Blick auf ein zukünftiges DCAT-AP.de geprüft und angepasst.

#### URI-Konzept als DCAT-AP.de Artefakt



CC BY SA International 4.0 „jinit“ AG und SID Sachsen für GovData“

Abbildung 4: Rolle des URI-Konzeptes im DCAT-AP.de Standard

Es wurden Regeln von anderen URI-Konzepten untersucht und durch eigene Regeln und Erfahrungen zur Wahrung von Interoperabilität und Zukunftssicherheit ergänzt. Die wesentlichen Grundlagen der Konventionen dieses Konzeptes stammen dabei aus den folgenden Quellen:

- a) W3C Best Practices [1],
- b) URI-Konzepte Großbritanniens [2] und
- c) URI-Konzepte Australiens [3].

Nach Überprüfung der Hostingtechnischen Möglichkeiten von DCAT-AP.de und GovData (Content Negotiation, Redirects, Rewrites,) wurden diese Konventionen mit der Fachgruppe und den Entwicklern des Standards DCAT-AP.de abgestimmt.

Werte, die in den Namensraum <http://dcat-ap.de> verweisen, etwa Datensätze unter der Deutschland Lizenz 2.0, werden künftig in Kommunen und anderen Verwaltungsebenen Datensatz für Datensatz hinterlegt.

Die DCAT-AP.de-Spezifikation greift auf Vorgaben dieses URI-Konzeptes zurück. Das URI-Konzept enthält Hinweise und Konventionen für die Verwendung von URIs bei den Datenbereitstellern von GovData und für die Bereitstellung von GovData-URIs.

## 2.2 DCAT-AP.de vs. GovData.de

Die Geschäfts- und Koordinierungsstelle GovData betreibt im Auftrag der beteiligten Partner der Kooperation [5] das Datenportal GovData sowie die dafür notwendige semantische Infrastruktur wie den Standard DCAT-AP.de. Zum Betrieb wurde die Einrichtung der URL „dcat-ap.de“ und weiterer darunterliegender Namensräume beschlossen. Unter anderem soll damit das Prinzip der Trennung von Metadatenstandard einerseits und Daten, Datenbeschreibung und Infrastruktur für GovData andererseits weiter vorangetrieben werden. Außerdem soll dem Prinzip der Vermeidung von Organisationsnamen in URIs folgend, denkbar nah am Standard selbst, eine URI im Namensraum dcat-ap.de zur Verfügung stehen und die Entwicklung des Standards DCAT-AP.de von der Entwicklung des Kooperationsverbundes GovData entkoppelt werden. Deshalb erfolgt in diesem Konzept die Festlegung von (Ontologie-)URIs für DCAT-AP.de und (Datensatz)-URIs für GovData.



Abbildung 5: Logo von GovData

## 2.3 Linked Data-Prinzipien

DCAT-AP.de schafft neue, zentral verwendbare Vokabulare für nationale Wertelisten und bietet mit diesem URI-Konzept den nachhaltigen Rahmen für zukünftige Änderungen an DCAT, DCAT-AP und DCAT-AP.de. DCAT-AP.de setzt dabei wie DCAT-AP als RDF-Vokabular auf Linked Data-Prinzipien gemäß dem 5-Sterne Reifegradmodell von Tim Berners-Lee:

- kein Stern: Daten im Web, Format unerheblich, ohne offene Lizenz
- 1 Stern: Daten im Web, Format unerheblich, mit offener Lizenz
- 2 Sterne: Daten in strukturiertem Format (z.B. Excel)
- 3 Sterne: Daten in strukturiertem, nicht proprietärem Format (z.B. CSV statt Excel)
- 4 Sterne: Verwendung von eindeutigen URLs, sodass Datensätze verlinkt werden können
- 5 Sterne: Verlinkung der eigenen Daten mit anderen Daten, um Kontext herzustellen



Abbildung 6: 5 Star Schema

DCAT-AP und DCAT-AP.de entsprechen dabei dem Umsetzungsgrad von 5 Sternen.

## 2.4 Arten von URIs

URIs können in vielerlei Gestalt auftreten:

1. URIs zur Identifikation

Beispiel: [www.example.com/id/alice](http://www.example.com/id/alice)

2. URIs zu Datenstrukturen

Beispiel: <https://www.govdata.de/web/guest/daten/-/details/b4dc51a9-ee0c-4ea1-9156-ccf21e234f94bkg>

3. URIs zu Distributionen

Beispiel: <http://www.geoportal.de/gds/xml.php?uuid=b4dc51a9-ee0c-4ea1-9156-ccf21e234f94>

4. URIs zu Massen-Download von Datenportalen (Bulk-URIs)

Beispiel: <http://www.govdata.de/bulk/>

5. URIs zu semantischen Konzepten, wie Codelisten, Taxonomien oder Ontologien (def)

Beispiel: <http://dcat-ap.de/def/plannedAvailability/>

6. Purl-URIs sind URIs zu semantischen Konzepten im Namensraum für persistent URLs (purl.org)

Beispiel: <http://purl.org/adms/publishertype/1.0>

7. Hash-URIs

Beispiel: <http://dcat-ap.de/def/plannedAvailability#available>

8. Slash-URIs

Beispiel: <http://dcat-ap.de/def/plannedAvailability/available>

9. Geo-URIs sind URIs, die geometrische Angaben (Punkte, Koordinatensystemtypen) in der URI tragen

Beispiel: `<geo:51.477811,-0.001475;u=0;crs=wgs84>----`

## 2.5 URIs als Bestandteil von DCAT-AP

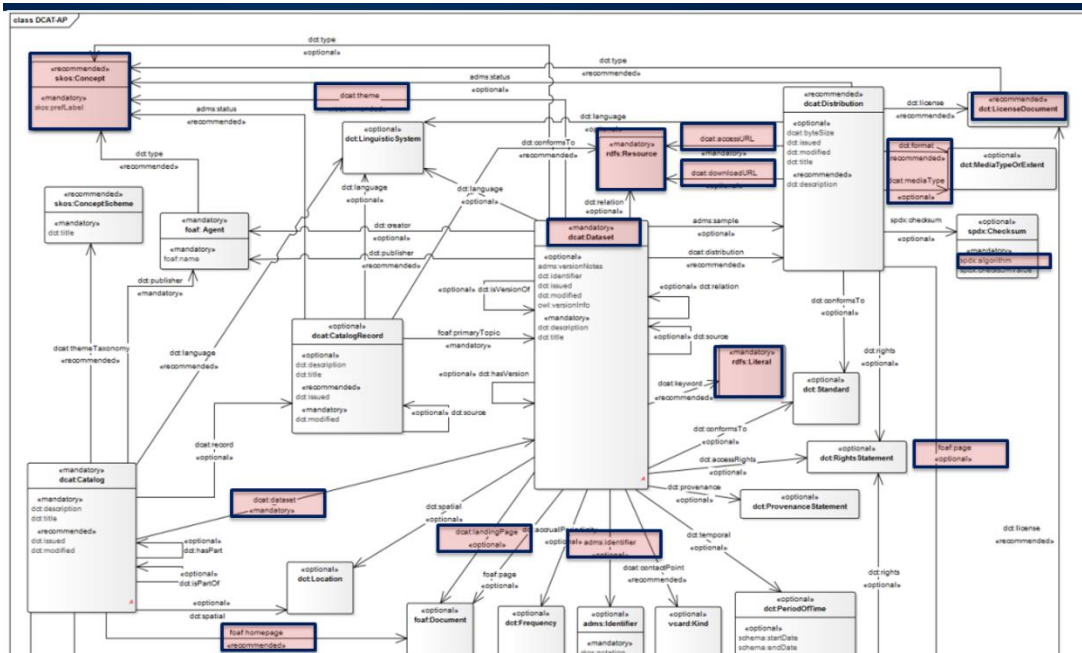


Abbildung 7: URIs in DCAT-AP

URLs und URIs in DCAT-AP sind zum Beispiel:

1. Sämtliche Eigenschaften die per `skos:concept` oder `rdfs:literal` auf URIs verweisen
2. `dcat:Catalog->foaf:homepage` (<http://xmlns.com/foaf/0.1/homepage>) mit der Beschreibung: „This property refers to a web page that acts as the main page for the Catalogue.“
3. `dcat:Distribution->dcat:accessURL` (<http://www.w3.org/ns/dcat#accessURL>) mit der Beschreibung: „This property contains a URL that gives access to a Distribution of the Dataset. The resource at the access URL may contain information about how to get the Dataset.“
4. `dcat:Distribution->dcat:downloadURL` (<http://www.w3.org/ns/dcat#downloadURL>) mit der Beschreibung: “This property contains a URL that is direct link to a downloadable file in a given format.“
5. `dct:identifier` (<http://purl.org/dc/terms/identifier>) mit der Beschreibung: „This property contains the main identifier for the Dataset, e.g. the URI or other unique identifier in the context of the Catalogue.“
6. `dcat:Dataset->dcat:landingPage` (<http://www.w3.org/ns/dcat#landingPage>) mit der Beschreibung: “This property refers to a web page that provides access to the Dataset, its Distributions and/or additional information.“
7. `dct:Document` ([http://xmlns.com/foaf/spec/#term\\_Document](http://xmlns.com/foaf/spec/#term_Document)) mit der Beschreibung: „A textual resource intended for human consumption that contains information, e.g. a web page about a Dataset.“

## 2.6 URIs als Bestandteil von DCAT-AP.de

URIs sind wichtige Bausteine für einen RDF-basierten Datenaustausch. Die DCAT-AP.de-Spezifikation schränkt zusammen mit dem Konventionenhandbuch die Verwendung von URIs weiter ein.

Damit URIs verlässlich zum Verweisen verwendet werden können, müssen diese Namensräume möglichst langlebig, d.h. persistent und nach Linked Data-Prinzipien ausgestaltet angelegt werden. Daher werden die für DCAT-AP.de benötigten Vokabulare im dcat-ap.de Namensraum angelegt.

Die URL der Liste der Datenbereitsteller an GovData.de befindet sich im Namensraum DCAT-AP.de:

<http://dcat-ap.de/def/contributors/deutscheNationalbibliothek>.

Die Liste aller im GovData-Informationsverbund akzeptierten Lizenzen wird durch die Geschäfts- und Koordinierungsstelle GovData gepflegt und unabhängig vom Releasezyklus der DCAT-AP.de Spezifikation aktualisiert. Gemäß Regel 25 erhält die Liste daher eine versionierte und eine unversionierte Verweismöglichkeit als URI.

Versionierte Liste der Lizenzen: <http://dcat-ap.de/def/licenses/1.0/>.

Aktuellste Liste der Lizenzen: <http://dcat-ap.de/def/licenses/>.

Soll also auf eine konkrete Lizenz-URI im GovData-Informationsverbund, z.B. der "Deutschland Lizenz 2.0 Namensnennung", in der aktuellen Version verwiesen werden, so ist folgende Slash-URI zu verwenden:

Konkreter Wert der Liste der Lizenzen: <http://dcat-ap.de/def/licenses/dl-de-by/2.0>.

Weitere als Codelisten verwendete Aufzählungen, die zum Zeitpunkt der Erstellung des DCAT-AP.de-Schemas von hoheitlicher Stelle gepflegt und nicht verlässlich in einem für RDF geeigneten URI-Format angeboten wurden, können im DCAT-AP.de URI-Raum genutzt werden.

Dies betrifft auch Wertelisten, die nicht „GovData-fachspezifisch“ sind, in der folgenden Aufstellung blau markiert.

Sollten zukünftig bessere Alternativen verfügbar sein, so wird für die Bestandsdaten mittels geeigneter (same:as) Mechanismen von zentraler DCAT-AP.de Stelle aus weiterverwiesen, ohne das Datenbereitsteller ihre Bestände dezentral korrigieren müssen.

Die Liste der im Juni 2017 geplanten dcat-ap.de URIs (blau = temporär) ist

1. Liste der Ebenen der geopolitischen Kodierung ([def/politicalGeocoding/level](#))
2. Liste der Regionalschlüssel und Bundesländer für den konkreten verwaltungspolitischen Geobezug als URI ([def/politicalGeocoding/regionalKey](#) und [def/politicalGeocoding/stateKey](#))
3. Liste der GovData Datenbereitsteller ([def/contributors](#) bei dcatde:contributorID)
4. Liste der Lizenzen ([def/licenses](#) bei dct:license bei dcat:Distribution)
5. Liste der Datenstrukturtypen ([def/datasetTypes](#) als dct:type bei dcat:Dataset)
6. Liste der zugesicherten Verfügbarkeiten ([def/plannedAvailability](#) als dcatde:plannedAvailability in dcat:Distribution)
7. Liste der Algorithmen ([def/hashAlgorithms](#) von spdx:algorithm in dct:Distribution)



## 2.7 URIs beim GovData Datenbereitsteller

Datenbereitsteller, die ihre offenen Daten mittels CKAN oder DCAT-AP oder DCAT-AP.de verwalten, haben eigene Angaben im URI-Format.

Beispiele von lokalen Distribution-URIs bei Datenbereitstellern:

1. Lizenztexte:

[http://www.geolizenz.org/modules/geolizenz/docs/1.2.1/GeoLizenz\\_V1.2.1\\_Open\\_082213\\_DE.pdf](http://www.geolizenz.org/modules/geolizenz/docs/1.2.1/GeoLizenz_V1.2.1_Open_082213_DE.pdf)

2. Eine Ressource im CSV-Format:

[http://www.bam.de/de/service/amtl\\_mitteilungen/sprengstoffrecht/sprengstoffrecht\\_medien/publi-idnr-sprengg.csv](http://www.bam.de/de/service/amtl_mitteilungen/sprengstoffrecht/sprengstoffrecht_medien/publi-idnr-sprengg.csv)

3. Ein Beispiel für eine schlechte URI eines Datenbereitstellers für eine Ressource im PDF-Format:

[https://www.datenhalter.de/SharedDocs/Downloads/DE/Publikationen/Statistik/2014/Standardtabelle/n/einzelstatistikkatalog\\_\\_pdf?&templateId=raw&property=publicationFile.pdf&sessionId=123456789](https://www.datenhalter.de/SharedDocs/Downloads/DE/Publikationen/Statistik/2014/Standardtabelle/n/einzelstatistikkatalog__pdf?&templateId=raw&property=publicationFile.pdf&sessionId=123456789)

Hier wird mit Dateiendung und Dateiformat doppelt im Namen und ohne Content Negotiation unter Technologiebezug (SharedDocs) mit überflüssigen, nicht-menschenlesbaren Informationen (templateID=raw), (property=publicationFile) auf eine durch Ablauf der SessionID nicht mehr existente Ressource verwiesen.

### 3 Übersicht über alle Regeln

Regel 1:	Nach dem 01.07.2017 neu angelegte URIs für offene Daten der öffentlichen Verwaltung in Deutschland SOLLEN eine Gültigkeitsdauer von mindestens 10 Jahre haben. ....	5
Regel 2:	HTTP-Links als URIs MÜSSEN nach Linked Data-Prinzipien verwendet werden, damit URIs aufgelöst werden können. Das bedeutet, dass im Internet zugängliche weitere Metadaten und Informationen (üblicherweise RDF und HTML) hinter den Adresslinks der Konzepte stehen müssen. ....	5
Regel 3:	Für URIs semantischer Konzepte unter DCAT-AP.de MUSS folgende Aufbauvorschrift angewendet werden: <code>http://dcat-ap.de/def/{konzept}</code> ; <code>konzept = Codeliste oder URI auf Datei in SKOS:notation</code> . ....	5
Regel 4:	Für URIs unter GovData und Portalen anderer Datenbereitsteller SOLL die folgende Aufbauvorschrift angewendet werden: <code>http://portal.de/{type}/{child-id}</code> ; <code>{type=def, dataset, bulk}</code> . ....	6
Regel 5:	Alle verfügbaren Repräsentationen SOLLEN durch dedizierte Erweiterungen der URI direkt zugänglich sein. ....	6
Regel 6:	In Fällen, in denen mehr als eine Repräsentations-URI angeboten wird, SOLL es auch möglich sein, mit Content Negotiation auf beide zuzugreifen. ....	6
Regel 7:	Bestandteile von URIs SOLLEN keine kurzlebigen Informationen enthalten, die ihre Auflösbarkeit zeitlich begrenzen. ....	6
Regel 8:	URIs für offene Datensätze SOLLEN technologieneutral formuliert sein und keine Bestandteile enthalten, die z.B. auf das Veröffentlichungssystem verweisen. ....	6
Regel 9:	URIs SOLLEN verständlich und menschenlesbar sein. ....	7
Regel 10:	Um DCAT-AP-Eigenschaften mit URIs zu benennen, SOLL auf Englisch zurückgegriffen werden, sofern es sich um einen sinnvoll ins Englische übersetzbaren Begriff handelt. ....	7
Regel 11:	Abgesehen von Personen der Zeitgeschichte MÜSSEN URIs anonymisiert sein und dürfen keine Personennamen enthalten. ....	7
Regel 12:	Bereits als ID bestehende andere persistente URIs SOLLEN nachgenutzt werden, wenn dem keine schwerwiegenden Gründe entgegenstehen. ....	7
Regel 13:	Es SOLL bei Wertelisten mit Releasezyklen außerhalb des DCAT-AP Standards für jede Version ein Versions-URI bereitgestellt werden, unter der diese erreichbar ist. ...	7
Regel 14:	Es SOLL eine nicht-versionierte URI bereitgestellt werden, welcher jeweils auf die aktuell gültige Version einer Werteliste verweist. Beispiel: <code>http://dcat-ap.de/def/licenses</code> . ....	7
Regel 15:	Der Singular SOLL verwendet werden, um ein bestimmtes physisches oder abstraktes „echtes Ding“ zu bezeichnen, also z.B. „/dataset/3434“. ....	8
Regel 16:	Der Plural SOLL angewendet werden, um eine Liste von Dingen zu benennen, also z.B. <code>http://dcat-ap.de/def/hashAlgorithms/</code> . ....	8

Regel 17:	Groß-/Kleinschreibung SOLL nicht als Unterscheidungsmerkmal zwischen URIs verwendet werden. ....	8
Regel 18:	URIs SOLLEN außerhalb des Domain-Namens in lowerCamelCase geschrieben sein. ....	8
Regel 19:	URIs zu Linked Data Datensätzen SOLLEN auf die Darstellung von Zeichen als Prozentdarstellung verzichten, sie SOLLEN also nur aus Ziffern, Buchstaben und den Sonderzeichen „-“, „.“, „_“ und „~“ bestehen. Die reservierten Zeichen SOLLEN entsprechend ihrer Bedeutung im Dokumentenpfad eingesetzt werden. (Siehe: <a href="https://de.wikipedia.org/wiki/URL-Encoding">https://de.wikipedia.org/wiki/URL-Encoding</a> ).....	8
Regel 20:	Lizenzen mit obligatorischer Namensnennung des Lizenzgebers SOLLEN den Bestandteil „-by“ im Lizenzcode führen. ....	9
Regel 21:	Lizenzen mit Zero-Kennung SOLLEN den Bestandteil „-zero“ im Lizenzcode führen... 9	9
Regel 22:	Lizenzen, die eine Bearbeitung der Daten untersagen, SOLLEN den Bestandteil „-nd“ im Lizenzcode führen. ....	9
Regel 23:	Lizenzen mit Share-Alike Attribution SOLLEN den Bestandteil „-sa“ im Lizenzcode führen. ....	9
Regel 24:	Lizenzen mit „Nicht für kommerzielle Nutzung“ Attribution SOLLEN den Bestandteil „-nc“ im Lizenzcode führen. ....	9
Regel 25:	Versionen von Lizenzen SOLLEN über die URI abgebildet werden.....	9
Regel 26:	Die Länderkennung des Rechtsraumes einer Lizenz (die Portierung in eine Sprache) SOLL über die URI abgebildet werden. ....	10
Regel 27:	Die Reihenfolge der generellen Aufbauvorschrift von Lizenz-URIs SOLL wie folgt umgesetzt werden: Lizenzakronym-zero-ByClause-NonCommercial-ShareAlike-NoDerivativeWork-Portierungsländercode/VersionMajor.VersionMinor. ....	10

## 4 Quellen

Analysierte Datenportale:

<https://www.govdata.de/>

<https://opendata.swiss/de/>

<https://www.europeandataportal.eu/>

<https://data.gov.uk/>

<https://data.overheid.nl/>

<http://opendata.gov.nl.ca/>

Analysierte URI-Konzepte:

[1] Niederlande: [https://www.w3.org/2013/04/odw/odw13\\_submission\\_14.pdf](https://www.w3.org/2013/04/odw/odw13_submission_14.pdf)

[2] UK: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/60975/designing-uri-sets-uk-public-sector.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60975/designing-uri-sets-uk-public-sector.pdf) bzw. <https://data.gov.uk/resources/uris>

[3] Australien: <https://github.com/AGLDWG/TR/wiki/URI-Guidelines-for-publishing-linked-datasets-on-data.gov.au-v0.1>

Analysierte Fachpublikationen:

[1] W3C und Share PSI Best Practices: <https://www.w3.org/2013/share-psi/bp/>

[2] Data Supplier Guidelines des European Open Data Portals:  
[https://www.europeandataportal.eu/sites/default/files/edp\\_s1\\_gdl\\_data-supplier-guidelines\\_v1.pdf](https://www.europeandataportal.eu/sites/default/files/edp_s1_gdl_data-supplier-guidelines_v1.pdf)

[3] DG Informatics Studie über Persistente URIs:  
<https://joinup.ec.europa.eu/sites/default/files/document/2013-02/D7.1.3%20-%20Study%20on%20persistent%20URIs.pdf>

[4] VOID-Vokabular: <http://vocab.deri.ie/void#uriSpace>

[5] RFC 3986: <http://www.ietf.org/rfc/rfc3986.txt>

[6] „Cool URIs don't change“: <https://www.w3.org/TR/cooluris/#r303gendocument>

[7] „Oasis Naming Guidelines“: <http://docs.oasis-open.org/specGuidelines/namingGuidelines/metadata.html#RDDLS>

[8] XML Coverpages über camelCase: <http://xml.coverpages.org/camelCase.html#un-cefact>

[9] Phil Archer zusammenfassend über URI-Persistenz:  
<http://philarcher.org/diary/2013/uripersistence/#itUriFormat>