

LEITFADEN FÜR AUSSAGEKRÄFTIGE METADATEN

Eine Veröffentlichung der mFUND-Begleitforschung des WIK

Februar 2019



Bundesministerium
für Verkehr und
digitale Infrastruktur



WIK-Begleitforschung



Wissenschaftliches Institut für
Infrastruktur und Kommunikationsdienste

Impressum

WIK Wissenschaftliches Institut für
Infrastruktur und Kommunikationsdienste GmbH
Rhöndorfer Str. 68
53604 Bad Honnef
Deutschland
Tel.: +49 2224 9225-0
Fax: +49 2224 9225-63
eMail: info(at)wik.org
www.wik.org

Vertretungs- und zeichnungsberechtigte Personen

Geschäftsführer und Direktor	Dr. Iris Henseler-Unger
Direktor	
Abteilungsleiter	
Post und Logistik	Alex Kalevi Dieke
Direktor	
Abteilungsleiter	
Netze und Kosten	Dr. Thomas Plückebaum
Leiter Verwaltung	Karl-Hubert Strüver
Vorsitzende des Aufsichtsrates	Dr. Daniela Brönstrup
Handelsregister	Amtsgericht Siegburg, HRB 7225
Steuer Nr.	222/5751/0722
Umsatzsteueridentifikations Nr.	DE 123 383 795

INHALTSVERZEICHNIS

1	EINLEITUNG	2
1.1	WARUM OPEN DATA?	2
1.2	DIE BEDEUTUNG AUSSAGEKRÄFTIGER METADATEN	
2	EMPFEHLUNGEN ZU METADATENFELDERN DER mCLOUD	3
2.1	LANGTITEL DES DATENSATZES	3
2.2	BESCHREIBUNG DES DATENSATZES	3
2.3	DATENANBIETER/-BEREITSTELLER	5
2.4	KATEGORIE	5
2.5	AKTUALITÄT DER DATEN	5
2.6	AKTUALITÄT DER DATENSATZBESCHREIBUNG	6
2.7	NUTZUNGSBEDINGUNG/LIZENZEN	6
2.8	NUTZUNGSHINWEISE	7
2.9	ART DES ZUGANGS	7
3	CHECKLISTE	8

1 EINLEITUNG

1.1 WARUM OPEN DATA?

Open Data bezeichnet den freien Zugang und die entgeltfreie Bereitstellung von maschinenlesbaren Daten. Werden die Daten ausschließlich von der öffentlichen Verwaltung bereitgestellt, spricht man auch von Open Government Data. Auch privatwirtschaftlich agierende Unternehmen stellen Daten offen zur Verfügung und verfolgen damit durchaus auch kommerzielle Interessen. Die Deutsche Bahn bspw. ist sehr aktiv in der Bereitstellung von Open Data. Ihr Ziel liegt dabei in der Suche nach neuen Anwendungsmöglichkeiten der Daten. Sie unterstützt diesen Prozess aktiv mit ihren [DB Hackathons](#). Hierbei entstehen u. a. Anwendungen, von denen sich die Deutsche Bahn eine Verbesserung ihres Service erwartet.

Die Bereitstellung von Open Data kann einen hohen gesellschaftlichen Mehrwert erzeugen: Der freie Zugang zu Daten kann Transparenz und Vertrauen in Politik und Verwaltung stärken sowie Innovationen und Wirtschaftswachstum ermöglichen. In Dänemark konnte durch die Veröffentlichung von Geodaten ein volkswirtschaftlicher Mehrwert von geschätzt 470 Mio. € im Jahre 2016 erwirtschaftet werden.¹ Studien bemessen das volkswirtschaftliche Potenzial von Open Data für Deutschland mit 12,1 Mrd. € bis 131,1 Mrd. € jährlich.²

Auf Grundlage von offenen Daten können innovative Anwendungen entwickelt werden, die zunächst einen Mehrwert für die Datenbereitsteller selbst, aber umso mehr auch für Bürgerinnen und Bürger schaffen können. Ein gutes Beispiel ist die App [Mein Moers](#), die frei verfügbare Informationen über die Stadt Moers bündelt und bereitstellt. Neue, prototypische Anwendungen werden regelmäßig auf kommunalen Hackathons, wie bspw. dem [Münsterhack](#), entwickelt. Für Wien wurden auf Basis der von der Stadt zur Verfügung gestellten Datensätze bereits [über 240 Apps](#) entwickelt. Das Spektrum reicht hier von Routenplanern für den ÖPNV und Informationen zu öffentlichen Leihrädern über Cityguides und Veranstaltungskalender bis hin zur Visualisierung von Pegelständen und Informationen zur Wasserqualität.

Open Data können unterschiedlichen Ursprungs und Natur sein. Sie können bspw. über Sensoren automatisch erhoben und mit geringem Zeitverzug über Open Data Plattformen bereitgestellt werden, denkbar bspw. bei Pegelständen von Gewässern. Oder sie werden manuell erhoben und eher selten aktualisiert. Ein Beispiel hierfür wäre die Liste aller Straßennamen einer Stadt.

Damit Open Data tatsächlich gefunden und genutzt werden können, ist eine Beschreibung der Daten mit aussagekräftigen Metadaten notwendig.

Mit diesem Leitfaden wollen wir die Bedeutung von Metadaten für die Bereitstellung von Open Data aufzeigen und Tipps geben, diese sinnvoll zu gestalten.

1 PWC (2017), The impact of the open geographical data – follow up study, <https://sdfe.dk/media/2917052/20170317-the-impact-of-the-open-geographical-data-management-summary-version-13-pwc-qrvkvdr.pdf>.

2 KAS (2016), Open Data. The Benefits. Das volkswirtschaftliche Potenzial für Deutschland, https://www.kas.de/c/document_library/get_file?uuid=3fbb9ec5-096c-076e-1cc4-473cd84784df&groupId=252038.

1.2 DIE BEDEUTUNG AUSSAGEKRÄFTIGER METADATEN

Metadaten sind strukturierte Informationen, die Merkmale von Daten beschreiben. Sie erleichtern es, relevante Daten in einer Vielzahl verfügbarer Datensätze in Datenportalen wie der mCLOUD (www.mcloud.de) zu suchen, zu finden und zu selektieren. Das Prinzip der Beschreibung mit Metadaten kennt man allgemein aus der Welt der Bibliotheken, in denen Bücher mit Metadaten beschrieben werden: Titel, Autor, Erscheinungsjahr und Klappentext sind Informationen, ohne die ein Buch in einem Bibliothekskatalog nicht gefunden werden kann. Aussagekräftige Metadaten stellen sicher, dass Datensätze von interessierten Nutzern gefunden werden und so ihr volles Potenzial entfalten können.

Struktur und Bedeutung der Metadaten unterschiedlicher Datensätze sollten möglichst einheitlich sein, damit sie auch über einen längeren Zeitraum und von verschiedenen Personen einfach genutzt werden können. Für die Beschreibung von Open Data bietet sich die Verwendung eines Standards wie DCAT-AP.de an, der deutschen Version eines europäischen Standards für Metadaten.³ Für den Anwendungsfall der mCLOUD wurden bestimmte Metadatenfelder ausgewählt, für den dieser Leitfaden im Folgenden Empfehlungen geben will.⁴

2 EMPFEHLUNGEN ZU METADATENFELDEN DER mCLOUD

2.1 LANGTITEL DES DATENSATZES

Der Langtitel beschreibt mit wenigen Worten, was der Inhalt des Datensatzes ist. Damit ist der Langtitel das Erste, was ein potenzieller Nutzer bei der Suche von Daten auf der mCLOUD findet. Er ist entscheidend dafür, dass Daten von Nutzern überhaupt gefunden werden können.

Der Titel sollte den Datensatz eindeutig identifizieren. Geltungsbereich der Daten, z. B. für eine Stadt, ein Bundesland oder bundesweit, und Zeitbezug sollten in den Titel aufgenommen werden, sofern diese Angaben sinnvoll möglich sind. Dadurch können Datennutzer schnell und einfach bewerten, ob Suchergebnisse für sie relevant sind.

Beispiel

Stadt Bonn: Bußgelder fließender Verkehr 2016

3 Eine Einführung zur Verwendung von DCAT.AP.de findet sich z. B. auf <https://lab.technologiestiftung-berlin.de/projects/opendata-metadata-dcat-ap/index.html>.

4 Für die Metadaten auf der mCLOUD ist DCAT-AP.de als technischer Standard implementiert. Darüber hinaus gibt es einzelne mCLOUD-spezifische Metadatenfelder, wie bspw. „Kategorie“, die nicht Teil des Standards sind.

2.2 BESCHREIBUNG DES DATENSATZES

Die Beschreibung enthält alle Informationen, die sich aus dem Langtitel noch nicht ergeben haben. Je ausführlicher der Datensatz hier beschrieben wird, desto eher weiß ein potenzieller Nutzer, ob er mit den Daten etwas anfangen kann.

Es sollten die Herkunft und die Art der Daten ersichtlich werden, noch bevor auf sie zugegriffen wird. Falls die Daten in Tabellenform keine leicht verständlichen Spaltennamen haben, so können diese hier erläutert werden. Bitte achten Sie darauf, zusätzlich zu einschlägigen Fachwörtern und Abkürzungen auch allgemein verständliche Begriffe zu verwenden, damit die Beschreibung auch von fachfremden Nutzern verstanden werden kann.

Beispiel

Dieser Datensatz umfasst die Lichtsignalanlagen/Ampeln (mit den Standorten der jeweiligen Steuergeräte) im Eigentum der Hanse- und Universitätsstadt Rostock, bewirtschaftet durch die Stadtwerke Rostock AG, mit Informationen zu Nummer, Bezeichnung und Knoten-Nummer. Die Ressourcen werden in der Regel im folgenden Zeitabstand aktualisiert: 7 Tage

Sofern es hilfreiche Dokumente (z. B. Standards) oder Webseiten gibt, die die Daten oder verwendete Fachwörter genauer beschreiben, sollten entsprechende Links in der Datensatzbeschreibung angegeben werden.

Beispiele

Deutscher Wetterdienst: [Datenformat KML Mosmix \(pdf\)](#)
Baustellendaten im MDM: [Standard Datex II \(pdf\)](#)
[Bathymetrische Daten](#)

2.3 DATENANBIETER/-BEREITSTELLER

Institution, Firma, Behörde etc., die die Daten erstellt bzw. bereitstellt oder über die Veröffentlichung entschieden hat. Dies können unterschiedliche Institutionen sein. Es sollte hier diejenige Institution genannt werden, die als Ansprechpartner fungieren kann.

Beispiel

Stadt Bonn

2.4 KATEGORIE⁵

Zuordnung der Daten zu einer oder mehrerer der folgenden Kategorien. Die Kategorien lehnen sich an die Zuständigkeitsbereiche des Bundesministeriums für Verkehr und digitale Infrastruktur an:

Kategorie		Erläuterung
Straßen		Alle Daten, die einen Bezug zum Straßenverkehr haben (z. B. motorisierter Individualverkehr, ÖPNV, Fußgängerverkehr, Radverkehr, Car-Sharing etc.).
Klima und Wetter		Alle Daten mit Bezug zu Wetterinformationen und Klimatologie.
Wasserstraßen und Gewässer		Alle Daten aus dem Bereich der Schifffahrt, hydrologische Daten sowie Umweltinformationen aus dem Gewässerbereich.
Bahn		Hierunter fallen alle Daten mit Bezug zum schienengebundenen Verkehr, also neben der Bahn auch Straßen- und U-Bahn, wie auch andere an eine feste Infrastruktur gebundene Bahnen, wie Seilbahnen.
Infrastruktur		Alle Daten, die Teile der digitalen sowie der Verkehrsnetzinfrastruktur beschreiben. Darunter fallen Daten zur eigentlichen Infrastruktur (z. B. Straßennetz, Breitbandnetz, WLAN-Standorte), aber auch zu zugehörigen Serviceeinrichtung (z. B. Parkplätze, Verkaufsstellen, Standorte, etc.).
Luft- und Raumfahrt		In diese Kategorie fallen alle Daten aus dem Bereich des Luftverkehrs, wie auch von Satelliten und von Luftfahrzeugen aufgenommenen Daten (z. B. Satellitenbilder, Orthofotos, etc.).

Die Zuordnung zu Kategorien hilft Nutzern der mCLOUD, gewünschte Daten zu finden. Zum einen kann in Kategorien gestöbert werden, zum anderen kann im Anschluss an eine Stichwortsuche nach Kategorien gefiltert werden.

2.5 AKTUALITÄT DER DATEN

Für welches Datum oder welchen Zeitraum hat der Datensatz eine Gültigkeit? Wurde er nur einmalig erstellt oder wird er ständig aktualisiert? Dies kann mCLOUD-Nutzern bei der Einschätzung helfen, ob der Datensatz für die vorgesehene Verwendung noch relevant ist.

Bei „statischen“ Daten sollte das Datum angegeben werden, an dem der Datensatz erstellt wurde. Bei „dynamischen“ Daten sollte das Datum angegeben werden, seit dem der Datensatz bereitgestellt wird.

Der Inhalt dieses Feldes ist ein Freitext, z. B.

- 01.01.2018
- 2017
- Februar 2017 bis April 2017
- Seit Januar 2016 vierteljährlich
- Wöchentlich

2.6 AKTUALITÄT DER DATENSATZBESCHREIBUNG

Datum, an dem die Datensatzbeschreibung (Abschnitt 2.2) eingetragen wurde. Es kann dabei helfen, die Aktualität der Beschreibung mit der Aktualität der Daten (Abschnitt 2.5) zu vergleichen. So kann es passieren, dass eine Beschreibung mit der Zeit veraltet während stetig neue Daten bereitgestellt werden, unter Umständen mit neuen Eigenschaften. Daher sollte die bereitstellende Organisation alte Datensatzbeschreibungen mit Hilfe dieses Datums prüfen, um die Beschreibung mit dem Datensatz konform zu halten.

2.7 NUTZUNGSBEDINGUNG/LIZENZEN

Lizenz, unter welcher der Datensatz bereitgestellt wird. Prinzipiell kann jede beliebige Lizenz gewählt werden. Da die mCLOUD jedoch den freien Zugang zu Daten und ihre Nutzung fördern möchte, wird empfohlen, die Lizenz so wenig restriktiv wie möglich/nötig zu wählen. Bundesbehörden können sich in dieser Frage von der zentralen [Beratungsstelle des Bundesverwaltungsamtes](#) beraten lassen.

Beispiele

[Datenlizenz Deutschland – Namensnennung – Version 2.0](#)

[Datenlizenz Deutschland – Zero – Version 2.0](#)

[Creative Commons Namensnennung 4.0 international](#)

[Open Data Commons Open Database License 1.0](#)

[Open Data Commons Public Domain Dedication and Licence 1.0](#)

[Open Data Commons Attribution Licence 1.0](#)

[Public Domain Mark 1.0](#)

2.8 NUTZUNGSHINWEISE

Freitextfeld für ergänzende Informationen zur Nutzung der Daten, die über die Lizenzangabe (Abschnitt 2.7) hinausgehen. Im Falle von Daten, die im Rahmen von Förderprojekten entstanden sind, können hier Informationen zum Fördermittelgeber angegeben werden.

2.9 ART DES ZUGANGS

Die mCLOUD verweist nur auf unmittelbar verfügbare Datensätze. Diese können vom Bereitsteller auf verschiedenen Wegen angeboten werden. Je mehr Zugangsmöglichkeiten für Daten angeboten werden, desto flexibler ist der Zugang für potenzielle Interessenten. Es sollten alle verfügbaren Links zu den Daten aufgelistet werden.

Beispiele für Zugangsarten:

- Dateidownload, z. B. direkt über den Browser. Z. B. oft für CSV-Dateien angeboten.
- Download per FTP (File Transfer Protocol).
- AtomFeed
- Datenportal: Falls ein weiteres Metadatenportal oder ein eigener Metadatenkatalog (neben der mCLOUD) existiert
- WMS-Dienst
- WMTS-Dienst
- WFS-Dienst
- WCS-Dienst
- SOS-Dienst
- API (Application Programming Interface bzw. Programmierschnittstelle)

Beispiel

Verweis auf einen Dateidownload unter

https://unfallatlas.statistikportal.de/app/UnfalldatenDownload/Unfallorte2016_LinRef_csv.zip

sowie der Link zum Portal <https://unfallatlas.statistikportal.de/>

3 CHECKLISTE

Vielen Dank für Ihr Interesse, Datensätze in der mCLOUD bereitzustellen. Das Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) stellt mit der mCLOUD einen zentralen Zugangspunkt zu allen offenen Daten seines Geschäftsbereiches zur Verfügung und öffnet das Portal auch für öffentliche und private Anbieter aus dem Mobilitätsbereich, die ihre Daten ebenfalls dort anbieten können. Die folgende Checkliste soll Ihnen dabei helfen, die von Ihnen bereitgestellten Daten so mit Metadaten zu versehen, dass potenzielle Nutzer über die Suche eine relevante Auswahl von Datensätzen angezeigt bekommen und auf Basis der Metadaten schnell entscheiden können, ob der jeweilige Datensatz tatsächlich allen Anforderungen entspricht.

Der bereitgestellte Datensatz ist mit einem aussagekräftigen Titel versehen.

Die Liste der Suchergebnisse kann über hunderte von Einträgen enthalten. Potenzielle Datennutzer wollen schnell die Datensätze auswählen, die mit hoher Wahrscheinlichkeit von Interesse sind. Dabei hilft ein aussagekräftiger Titel, der alle relevanten übergeordneten Informationen enthält. Die Länge des Titels sollte ca. 10 Worte nicht überschreiten. Ein aussagekräftiger Titel sollte detailliert genug sein, um eine schnelle Beurteilung der Relevanz zu ermöglichen. Bspw. könnte der Titel „Kraftfahrzeuge in Bremen“ durch die Ergänzung „Kraftfahrzeuge in Bremen mit amtlichem Kennzeichen“ genauer beschrieben werden, um einem potenziellen Nutzer zu ermöglichen diesen Datensatz schnell auszuschließen, wenn an Zahlen von allen Kraftfahrzeugen, d. h. auch stillgelegten Fahrzeugen, Interesse besteht.

Der Titel oder die Beschreibung des Datensatzes enthält alle relevanten Suchterme.

Um einen Datensatz in der mCLOUD finden zu können, ist es wichtig, dass die Daten auch über die eingegebenen Suchbegriffe der potenziellen Nutzer gefunden werden können. Die Suchbegriffe werden dabei mit dem Titel oder der Beschreibung des Datensatzes in der mCLOUD abgeglichen. Achten Sie darauf, dass Ihre Beschreibung trotz spezifischer Fachwörter allgemein verständlich bleibt und dass neben den üblichen Abkürzungen auch die Langbeschreibung verwendet wird; so ist z. B. der „Hamburger Verkehrsverbund“ auch als „HVV“ bekannt. Die Frage, welche Suchbegriffe potenzielle Nutzer wahrscheinlich verwenden, ist nicht immer einfach zu beantworten. Es kann sehr hilfreich sein, Kolleginnen oder Kollegen, die nicht mit dem Datensatz vertraut sind, zu bitten, nach Ihrem Datensatz zu suchen. Nach welchen Begriffen suchen sie?

Messauftrag, Datenqualität, Messgeräte, Methodik etc. sind in der Beschreibung angegeben.

Verschiedene zusätzliche Faktoren haben einen Einfluss darauf, wie die Daten zu bewerten sind und ob sie der geplanten Datennutzung zugeführt werden können. So kann z. B. der verwendete Sensor Aufschluss darüber geben, welcher Messbereich abgedeckt wird. Eine Angabe, wie z. B. „die Messwerte wurden mit dem Sensor AZF43 der Firma Mustermann aufgenommen“ ermöglicht es potenziellen Nutzern, aus der Spezifikation des Sensors die für sie entscheidenden Daten zu entnehmen. Die Methodik der Datenerhebung kann durch einen kurzen Hinweis auf das Verfahren („nach XYZ-Verfahren erhoben“) genauer beschrieben werden. Auch die Angabe des ursprünglichen Messauftrags kann hilfreich sein, um zu entscheiden, ob die Daten für den geplanten Einsatz geeignet sind, z. B. durch die Angabe „Die Daten wurden im Rahmen der Einkommens- und Verbrauchsstichprobe des statistischen Bundesamtes erhoben“.

Der Datensatz ist einer relevanten Kategorie zugeordnet.

Gegenwärtig wird jeder Datensatz in der mCLOUD mindestens einer von sechs Kategorien zugeordnet, wie z. B. Straße, Infrastruktur, Klima und Wetter. Wenn Sie nicht sicher sind, welcher Kategorie die Daten zugeordnet werden sollten, kann es hilfreich sein, (gerne auch fachfremde) Kolleginnen oder Kollegen um Rat zu bitten: Welcher Kategorie würden sie den Datensatz zuordnen?

Die Aktualität der Datensatzbeschreibung entspricht dem angegebenen Zeitraum.

Bei einem Update eines Datensatzes ist darauf zu achten, dass die Datensatzbeschreibung weiterhin aktuell ist und dem angegebenen Datum entspricht, oder ob die Datensatzbeschreibung angepasst werden muss.

Die angegebene Lizenz ermöglicht möglichst vielen potenziellen Nutzern, die Daten zu verwenden.

Um Ihren Datensatz einer möglichst großen Gruppe von Nutzern zur Verfügung zu stellen, schränken Sie dessen Nutzung nicht durch eine unnötig restriktive Lizenz ein. Bitte prüfen Sie intern, ob es möglich ist, eine freie Lizenz zu verwenden, um einen großen Nutzerkreis zu ermöglichen.

Dieser Leitfaden entstand im Arbeitsforum Standardisierung/mCLOUD der mFUND-Begleitforschung des WIK. Autoren des Leitfadens sind Sebastian Goerke, Dr. Holmer Hensen, Niklas Petersen, Dr. Till Riedel, Tim Ruß und Julia Wielgosch. Das Arbeitsforum bietet den im mFUND geförderten Projekten die Gelegenheit, Erfahrungen mit Standards und Datenportalen auszutauschen und Empfehlungen zur Standardisierung zu erarbeiten, insbesondere hinsichtlich des Datenportals mCLOUD. Mehr Informationen zum Arbeitsforum unter mfund.wik.org

Im Rahmen der **Forschungsinitiative mFUND** fördert das BMVI seit 2016 Forschungs- und Entwicklungsprojekte rund um digitale datenbasierte Anwendungen für die Mobilität 4.0. Mehr Informationen unter www.bmvi.de/mfund



Die **mFUND-Begleitforschung des WIK** unterstützt die effiziente und effektive Umsetzung des Förderprogramms. Mehr Informationen unter mfund.wik.org und [@WIKnews](https://twitter.com/WIKnews)

