

HANDREICHUNG METADATENQUALITÄT

Metadaten sind Daten, die Daten beschreiben. Was zunächst abstrakt klingt, lässt sich leicht an einem praktischen Beispiel verdeutlichen: Die meisten von uns hören gerne Musik. Im digitalen Umfeld, etwa beim Musik-Streaming, werden Musikdateien üblicherweise durch Metadaten ergänzt. Das heißt, die Datei eines Liedes enthält nicht nur die Tonspur, sondern auch Daten, die diese Tonspur beschreiben: Liedtitel, Albumtitel, Künstlername und weitere Metadaten.

Anhand dieser beschreibenden Daten können wir bequem aus Millionen von Titeln die Musik aussuchen, die wir hören möchten. So suchen wir beispielsweise nach dem Namen einer Sängerin und erhalten eine Übersicht ihrer verfügbaren Werke.

Dieses Prinzip gilt ebenso für offene Daten: Mehr als 30.000 Datensätze aus ganz Rheinland-Pfalz sind durch Metadaten auf der [Transparenz- und Offenen-Daten-Plattform](#) des Landes repräsentiert und

auffindbar. Anhand dieser Metadaten können Nutzende diese Datenmenge durchsuchen und für ihre Zwecke optimal nutzen.

Über die grafische Oberfläche der Plattform stehen zahlreiche Such- und Filtermöglichkeiten für die Nutzenden zur Verfügung. Darüber hinaus kann der Metadatenkatalog über eine dokumentierte Programmierschnittstelle (ein sogenanntes Application Programming Interface, API) abgerufen und durchsucht werden.

Entscheidend ist jedoch nicht das alleinige Vorhandensein der Metadaten, sondern ihre Qualität. Nur mit „sprechenden“ Metadaten lassen sich die Daten gut finden. Die Verantwortung für die Qualität der Metadaten liegt dabei bei den Datenbereitstellenden.

Diese Handreichung unterstützt Sie als Datenbereitstellende dabei, vollständige und hochwertige Metadaten für Ihre Daten zu erstellen.

Der Metadatensatz im Fokus

Im Deutschen Metadatenstandard [DCAT-AP.de](#) sind für viele Eigenschaften Wertebereiche über vordefinierte Vokabulare festgelegt. Bei diesen Metadaten ist genau vorgegeben, welche Inhalte in die Metadatenfelder eingetragen werden dürfen.

Eigenschaften, die frei befüllt werden können, bieten dagegen mehr Spielraum zur Erhöhung der Metadatenqualität. Auf diese konzentriert sich diese Handreichung folglich.

Titel (dct:title)

Der Titel eines Datensatzes springt den Nutzenden der **Plattform** sofort ins Auge. Er ist damit eines der prominentesten Metadatenfelder und muss präzise über den Dateninhalt informieren.

Ein guter Titel enthält die folgenden Informationen:

- Eine prägnante Bezeichnung des Datensatzes ohne Abkürzungen.
- Eine eindeutige zeitliche Einordnung.
- Eine sprechende Angabe des geographischen Raumes bei Daten mit geographischem Bezug.

Beschreibung (dct:description)

Die Beschreibung ergänzt den Titel mit tieferem Einblick in den Dateninhalt. Sie sollte nicht den Titel wiederholen, sondern zusätzliche Informationen bieten, die die Auffindbarkeit des Datensatzes erhöhen.

Kriterien für eine gute Beschreibung

Urheber

Wer hat die Daten erfasst?

Methodik

Wie wurden die Daten erhoben?

Struktur

Wie sind die Daten aufgebaut?

Erhebungszeitraum

Wann wurden die Daten erhoben?

Geografische Zuordnung

Wo beziehungsweise für welches Gebiet wurden die Daten erhoben?

Qualitätsangaben

Welche Einschränkungen müssen für eine Auswertung bekannt sein?

/open

Diese Kriterien ermöglichen den Nutzenden die Einordnung und kritische Auseinandersetzung mit den Daten. Das folgende Beispiel für den [Datensatz zu Kreisstraßen in Rheinland-Pfalz](#) verdeutlicht den Qualitätsunterschied bei Anwendung der genannten Prinzipien:

Titel	Negativbeispiel	Positivbeispiel
Beschreibung	Kreisstraßen	Kreisstraßen in Rheinland-Pfalz
	Kreisstraßen	Kreisstraßen in Rheinland-Pfalz, die durch den Landesbetrieb Mobilität erfasst werden. Die Daten werden in regelmäßigen Abständen aktualisiert.

Schlagworte (dcat:keyword)

Schlagworte verbessern die Auffindbarkeit von Datensätzen, indem sie diese mittels allgemeiner Begriffe kategorisieren. Dadurch können Datensätze zu gleichen und ähnlichen Inhalten ausfindig gemacht werden.

Die Vergabe von Schlagworten sollte also nicht einzelfallspezifisch erfolgen. Ein Schlagwort, das lediglich für einen Datensatz gilt, ist wenig hilfreich. Vielmehr sollten im Vorfeld der Vergabe der Metadaten allgemeine Schlagworte definiert werden.

Diese ermöglichen die Zuordnung thematisch verwandter Datensätze. Im optimalen Fall orientiert sich die Auswahl an bestehenden Schlagworten auf der [Transparenz- und Offene-Daten-Plattform des Landes Rheinland-Pfalz](#) sowie dem Datenportal für Deutschland [GovData](#).

Wir empfehlen etwa drei bis fünf Schlagworte pro Datensatz auf open.rlp.de zu verwenden.

Gute und schlechte Praktiken für Verschlagwortung

Thema	Gute Praktiken	Schlechte Praktiken
Länge & Aufbau	Schlagwort besteht nur aus einem Wort (Ausnahme: Eigennamen) Beispiel: <i>Umwelt, Bundesministerium für Gesundheit</i> (Eigennamen)	Schlagworte aus mehreren Wörtern ohne Eigennamen Beispiel: <i>öffentliche Verwaltung Daten</i>
Inhaltliche Prägnanz	Schlagworte geben prägnant den Inhalt des Datensatzes wieder Beispiel: <i>Luftqualität, Kataster</i>	Schlagworte sind vage, unspezifisch oder beschreiben nicht Inhalt Beispiel: <i>Allgemein, Dokument, Eintrag</i>
Zulässige Abkürzungen	Abkürzungen verwenden, wenn amtlich oder allgemein gebräuchlich Beispiel: <i>EU, BMI, GIS</i>	Erfundene oder verwaltungsinterne Abkürzungen ohne breite Bekanntheit Beispiel: <i>AGTNBG</i> (unbekannt)
Abkürzungen ergänzen	Zusätzlich bestimmende Eigenschaften von Abkürzungen in Langform angeben Beispiel: <i>AGTierNebG, tierische Nebenprodukte</i>	Nur Abkürzung als Schlagwort verwenden Beispiel: <i>AGTierNebG</i>
Keine Zahlen	Zahlen nicht als Schlagworte verwenden	Jahreszahlen, Kennziffern oder reine Zahlen als Schlagworte

/open

Redundanzen vermeiden

Keine Dopplung von Bezeichnungen, die bereits in anderen Feldern vorhanden sind

Beispiel: *Umweltministerium im Feld „Veröffentlichende Stelle“, Schlagworte: Förderprogramme, Haushalt*

Doppelungen, z. B. Veröffentlichende Stelle auch als Schlagwort

Beispiel: *Umweltministerium* zusätzlich zum Feld „Veröffentlichende Stelle“

Synonyme verwenden

Synonyme ausdrücklich verwenden, um Suchtreffer zu verbessern

Beispiel: *Abfall, Müll*

Nur ein Begriff ohne Synonyme

Beispiel: *Abfall*

Sonderzeichen & Länge

Erlaubt sind alphanumerische Zeichen sowie „.“, „-“, „_“; max. 100 Zeichen

Beispiel: *CO2-Emissionen, Geo_Daten, E-Government*

Andere Sonderzeichen verwenden; Schlagworte > 100 Zeichen

Beispiel: *Ministerium Für Arbeit Soziales Transformation Und Digitalisierung Rheinland-Pfalz Aktuelle Legislaturperiode*

Herausgeber (dct:publisher)

Die Pflege der Daten liegt in der Verantwortung der Liefersysteme. Diese müssen daher immer eine E-Mail für die Kontaktaufnahme hinterlegen. Damit können Fehler in den Daten entweder das Liefersystem selbst oder die verantwortliche Stelle erreichen. Hierbei ist aber zu beachten, dass es

sich um eine allgemeine Poststelle handeln muss. Von der Angabe personenbezogener Daten (zum Beispiel E-Mailadressen von Sachbearbeitenden) ist – auch in Hinblick auf die Datenschutz-Grundverordnung (DSGVO) – abzusehen.

Weiterführende Informationen

Vorgaben zu den Metadaten ergeben sich aus [DCAT-AP.de](https://www.dcat-ap.de), dem deutschen Metadatenstandard zum Austausch offener Verwaltungsdaten. Anforderungen, die die Metadatenqualität betreffen, ergeben sich insbesondere aus dem [DCAT-AP.de Konventionenhandbuch](#).



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