



Anforderung an die Daten – Leitfaden –



Inhaltsverzeichnis

Einleitung.....	3
1. Datenscreening.....	3
1.1. Auswertung	5
2. Dateiformate wählen.....	6
3. Daten an GovData.de anbinden - Metadaten an GovData.de übertragen.....	6
3.1. Metadaten erfassen.....	6
4. Mit Metadaten die Daten auffindbar machen.....	7
5. Freie Lizenzen.....	8
5.1. Wie lässt sich die richtige Lizenz finden?	8
5.2. Datenlizenz Deutschland.....	8
5.3. Creative Commons Lizenzen.....	9
6. Wo erhalte ich Support?.....	9
7. Anhang.....	10

Version 1.6
August 2020

Impressum

Bundesverwaltungsamt

Kompetenzzentrum Open Data (CCOD)

50728 Köln

Telefon: 022899358-4842

E-Mail: OpenData@bva.bund.de

Internet: www.opendata.bund.de

Einleitung

Dieser Leitfaden unterstützt bei der:

- Identifikation geeigneter Datensätze
- Auswahl des Datei-Formates
- rechtlichen Bewertung
- Metadaten sowie
- Wahl der Lizenz

Wie sehen formale, rechtliche und technische Anforderungen an die bereitzustellenden Roh- und Metadaten aus? Welche Konventionen sind einzuhalten, was kann, muss oder soll nach bestimmten Regeln erfolgen und welche Formate sind dabei zulässig?

Sind einmal Open Data fähige Daten gefunden¹ (siehe dazu das Handbuch Kapitel 3.1.1) gilt es diese nach rechtlichen Kriterien zu filtern und anschließend die veröffentlichungsfähigen Daten in ein offenes Format zu überführen.

Danach gilt es die (Roh-)Daten auf GovData.de, dem zentralen Datenportal für Bund und Länder, auffindbar zu machen. Dies geschieht durch beschreibende Metadaten, die an GovData.de übermittelt werden.

Der Leitfaden dient als Anleitung und Erfassungshilfe. Er bildet eine Ergänzung zum Open Data Handbuch und wird somit als ein unterstützendes Papier veröffentlicht. Aussagen daraus können nicht im Rahmen gesetzlicher Verpflichtungen und deren Umsetzung geltend gemacht werden.

1. Datenscreening

Zu Beginn gilt es zu ermitteln, welche Daten seit dem 13.07.2017 erstmalig erstellt bzw. aktualisiert wurden.

Diese Dateninventur kann durch eine manuelle Ermittlung erfolgen. Zum Beispiel durch eine Abfrage in den Fachabteilungen, welche Daten dort im eigentlichen Auftrag der Behörde erstellt bzw. aktualisiert wurden.

Alternativ kann die Webseite Ihrer Behörde nach Daten durchsucht werden. Die einfachste aber ungenaueste Möglichkeit bietet die Suche mit Google-Suchoperatoren. Hierzu gibt man im Google-Suchfenster, die Webseite der Behörde und den gesuchten Dokumententyp an und startet die Suche. Als Ergebnis werden alle Dokumente im Internet angezeigt, die zu diesen Kriterien gefunden werden konnten.

¹ Vgl. Open-Data-Handbuch, Kapitel 3.1.1.

Beispiel:

Site:bva.bund.de filetype:csv



site:www.bva.bund.de filetype:csv



In diesem Beispiel erhält man mit der Abfrage alle veröffentlichten CSV-Dateien auf der Website des Bundesverwaltungsamtes.

Neben Dokumententypen wie CSV oder XML kann es interessant sein, auch nach PDF-Dokumenten zu suchen. Zwar sind PDFs in der Regel nicht maschinenlesbar, häufig lassen sich darin jedoch Übersichten oder Tabellen finden, die wiederum als Rohdaten zur Verfügung gestellt werden könnten.

Anschließend empfiehlt es sich, die Ergebnisse zu strukturieren. Für eine bessere Übersicht pflegen Sie die Angaben zu den Daten am Besten in einen Datenkatalog ein. Dieser verschafft Ihnen zudem einen Überblick über alle in Ihrem Haus erstellten Daten und kann bei weiteren Projekten nützlich sein.

Wählen Sie anschließend die Daten aus, die veröffentlichungsfähig sind und priorisieren Sie diese für die Veröffentlichung.

Nachfolgende Checkliste dient als Vorschlag zur internen Priorisierung von Daten.

Kriterium	Erläuterung	Bewertung (Punkte 0-5)
Geheimhaltung/ rechtliche Hindernisse	Unterliegen die Daten Geheimhaltungspflichten oder sonstigen rechtlichen Beschränkungen bzw. handelt es sich um infrastrukturkritische Daten?	0: Geheimhaltungspflicht gegeben 1: Einschränkungen vorhanden, änderbar 2: Einschränkungen vorhanden, sehr leicht änderbar (z. B: interne Richtlinien, Verwaltungskultur) 3: keine Einschränkungen
Personenbezug	Handelt es sich um personenbezogene Daten bzw. lassen sich Rückschlüsse auf Personen daraus ableiten?	0: Personenbezogene Daten 1: Zustimmung zur Veröffentlichung vorhanden (z. B. Förderdaten) 2: Anonymisierbare Daten 3: Kein Rückschluss auf Personen ableitbar, bzw. keine Verletzung schutzwürdiger Geheimhaltungsinteressen (§5 IFG)
Unternehmensbezug	Handelt es sich um unternehmensbezogene Daten bzw. lassen sich Rückschlüsse auf Unternehmen daraus ableiten?	0: Rückschluss auf Betriebs- oder Geschäftsgeheimnisse? 1: Einwilligung des Betroffenen liegt vor 2: Kein Rückschluss auf Betriebs- und Geschäftsgeheimnisse (§6 IFG)
Nutzungsrecht	Besitzt die Verwaltung das alleinige Nutzungsrecht der Daten?	0: Fehlendes Nutzungsrecht: Veröffentlichung ist nicht möglich 1: Lizenzkosten fallen an, Genehmigungen sind einzuholen 2: Lizenzkosten fallen an, Genehmigungen sind vorhanden

		3: Keine Lizenzkosten, aber Genehmigungen sind einzuholen 4: Keine Lizenzkosten, Genehmigungen vorhanden 5: Alleiniges Nutzungsrecht sichergestellt
Nutzen	Wie hoch wird der Nutzen für alle Zielgruppen eingeschätzt?	1: Der Nutzen ist gering 2: Der Nutzen ist durchschnittlich 3: Der Nutzen ist hoch
Aufwand	Wie hoch ist der Aufwand für die Veröffentlichung?	1: Der Aufwand ist hoch 2: Der Aufwand ist durchschnittlich 3: Der Aufwand ist gering
Inhaltliche Datenqualität	Wie hoch wird die Datenqualität eingeschätzt? (Vollständigkeit, zeitliche Nähe, Genauigkeit, Fehlerhaftigkeit,...)	1: Die Datenqualität ist gering 2: Die Datenqualität ist durchschnittlich 3: Die Datenqualität ist hoch
Technische Verfügbarkeit	Verfügbare Datenformate und Datenquellen, Maschinenlesbarkeit, offene Standards...	0: Daten sind nicht elektronisch verfügbar 1: Daten sind elektronisch verfügbar, jedoch nicht maschinenlesbar 2: Daten sind in maschinenlesbarem Format verfügbar 3: Daten sind leicht in ein offenes Format zu exportieren 4: Daten liegen in offenem Format vor

1.1. Auswertung

Die Daten mit einem hohen Punktwert sollten zuerst veröffentlicht werden bzw. können bedenkenlos veröffentlicht werden. Je nach Reduktion der Punkte sind eventuell Anpassungen notwendig.

Sollten einzelne Kriterien mit 0 bewertet werden (rot gekennzeichnet), ist dies Ausschlussgrund für eine Veröffentlichung. Kriterien, bei denen 0 nicht vorkommt, können zwar kumuliert ebenso wenige Punkte erhalten, dies stellt allerdings keinen Ausschlussgrund dar. Daten, die gelb gekennzeichnete Bewertungszahlen erhalten, müssen gemäß EGovG nicht veröffentlicht werden. Diese Daten können aber ggf. nach entsprechender Bearbeitung bereitgestellt werden, da hier ebenfalls ein Nutzen für andere gegeben sein kann.

Falls im ersten Schritt, aufgrund einer zu geringen Bewertung, von einer Veröffentlichung abgesehen wird, sollten Maßnahmen zur Erhöhung der Datenqualität getroffen werden. Im Anschluss kann eine erneute Bewertung durchgeführt werden.

Die fehlende Gewichtung der Kriterien impliziert einen gleichen Stellenwert jedes Kriteriums. Dies ist eine vereinfachende Annahme, die allerdings für das interne Datenscreening ausreicht. Es handelt sich um eine Einschätzung, die dabei helfen soll, veröffentlichungsfähige Datensätze zu identifizieren. Es ist allerdings auch möglich, den Kriterien unterschiedliche Gewichtungen zu verleihen.²

² in Anlehnung an: Krabina, Bernhard (KDZ), Lutz, Brigitte (Stadt Wien): Open-Government-Vorgehensmodell. Umsetzung von Open Government Version 3.0. Wien (2016), S25f. <https://kdz.eu/de/file/18417/download> (Stand 25.06.2020).

2. Dateiformate wählen

Daten, die als Open Data bereitgestellt werden, sollen in der Regel automatisiert von Maschinen ausgelesen und verarbeitet werden können. Damit Daten problemlos maschinenlesbar sind, müssen sie in einem offenen, strukturierten Format vorliegen. Ein offenes Format liegt vor, wenn kein Spezialprogramm benötigt wird, um die Informationen dieser Datei verarbeiten zu können. Eine Excel-Datei ist demnach zwar ein strukturiertes, jedoch kein offenes Dateiformat, da ein Spezialprogramm benötigt wird, um die Datei zu öffnen. Dies schränkt den Kreis der möglichen Nutzer ein. Kommage-trennte Tabellen, sogenannte CSV-Dateien, benötigen hingegen kein bestimmtes Programm, um betrachtet und bearbeitet werden zu können.

Um eine möglichst diskriminierungsfreie Weiterverwendung der Daten zu ermöglichen sollte deshalb auf offene und strukturierte Dateiformate gesetzt werden.

Folgende Übersicht listet die zurzeit für Open Data am besten geeigneten Formate auf:

- Dateiformate: XML, RDF, CSV, JSON, RDF, HTML

Diese Übersicht ist nicht in Stein gemeißelt. So können neue Programme oder Veränderungen an bestehenden Programmen dazu führen, dass diese Liste angepasst werden muss.

3. Daten an GovData.de anbinden - Metadaten an GovData.de übertragen

Die bereitzustellenden Daten werden über Metadaten auf GovData.de auffindbar gemacht. Dies erfolgt einmalig außer die Daten werden aktualisiert oder in sonstiger Weise angepasst. Dann sollte auch eine Aktualisierung der Metadaten erfolgen.

Um Daten als Datenbereinsteller an GovData.de anzubinden, benötigen Sie Zugang zum Bereitsteller-konto Ihres Ministeriums.

Zurzeit ist es nicht angedacht, dass jede*r Mitarbeiter*in, ein eigenes Konto auf GovData.de erhält. Eine entsprechende Koordination des Zugangs wird seitens des Open-Data-Beauftragten des Ministeriums sichergestellt.

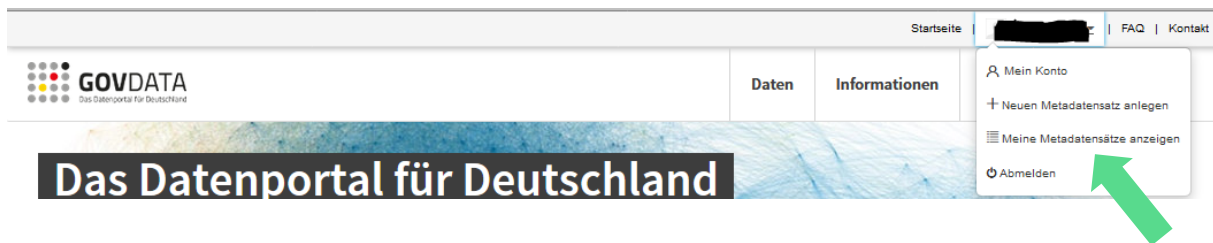
3.1. Metadaten erfassen

Für die Übertragung der Metadaten steht Ihnen zurzeit folgender Weg zur Verfügung:

1. (Manuelle) Bereitstellung per Webformular

Diese Methode ist ideal, wenn nur wenige Datensätze bereitgestellt werden. Die erforderlichen Eingabeprozesse sind leicht beherrschbar, so dass der Zeitaufwand überschaubar bleibt.

Dazu gehen Sie nach erfolgter Anmeldung auf Ihren Benutzernamen. Es öffnet sich ein Drop Down Feld. Klicken Sie nun auf „Neuen Metadatensatz anlegen“.



In einem neuen Fenster öffnet sich nun die Eingabemaske für die Eingabe von Metadaten über das Webformular.

4. Mit Metadaten die Daten auffindbar machen

Damit Ihre Daten auf dem GovData.de-Portal überhaupt gefunden werden können bedarf es beschreibender Metadaten. Diese Daten beinhalten unter anderen Informationen über die veröffentlichende Stelle, die Aktualität der Daten und verweisen auf den eigentlichen Datensatz.

Metadaten enthalten Information über die Merkmale von Datensätzen aber nicht die Daten selbst. Sie helfen, ihren Datensatz einzugliedern, wie Registerdaten zu einer Akte. Um Daten auf GovData.de auffindbar zu machen, ist es notwendig, die Metadaten nach einem einheitlichen Schema einzupflegen. Nur wenn Felder gleich benutzt werden und Mindestangaben gepflegt werden, können Ihre Daten auch gefunden werden.

Wichtig:

Auf GovData.de werden ausschließlich Metadaten erfasst. Die Daten selbst verbleiben bei Ihnen.

Der deutsche Metadatenstandard „DCAT-AP.de“ ist mit dem europäischen Metadatenstandard „DCAT-AP“ kompatibel, wodurch eine Datenweitergabe „nach Oben“ zum [Open-Data-Portal der Europäischen Union](#) gewährleistet ist. Die Metadatenstruktur von GovData.de vom IT-Planungsrat als verbindliche Nutzung des Interoperabilitätsstandards DCAT-AP beschlossen.

Unter **Metadaten** versteht man strukturierte Daten, die Informationen über andere Informationsressourcen enthalten. Metadaten beschreiben also den eigentlichen Datensatz. Sie werden erforderlich, wenn es größere Datenmengen zu verwalten gibt. Ein Merkmal von Metadaten ist, dass sie maschinell lesbar und auswertbar sind.

Mögliche Metadaten bei einem Buch wären zum Beispiel:

- Autor,
- Titel,
- Verlag,
- Erscheinungsjahr,
- Anzahl der Seiten,
- ISB Nummer

Metadaten sind der Schlüssel, wenn es darum geht, Datensätze nutzbar zu machen. Ohne sie sind Ihre Daten zwar veröffentlicht aber für niemanden auffindbar. Bei den Verweisen handelt es sich zumeist um Weblinks, wodurch eine Ressource eindeutig identifizierbar ist. Die Gefahr Doubletten anzufertigen wird damit minimiert.

Die beschreibenden Metadaten liefern die nötigen Informationen, um den Inhalt des Datensatzes darzustellen, ohne den Datensatz selbst öffnen zu müssen. Metadatenstandards vereinheitlichen die auszufüllenden Felder, um ein einheitliches Qualitätsniveau zu erreichen. Neben obligatorischen Feldern, wie Angaben zum Datenbereiter und den Titel des Datensatzes, können auch empfohlene und optionale weitere Metainformationen im neuen Metadatenchema beigefügt werden. Dies ermöglicht eine noch präzisere Beschreibung der Daten und erhöht die Chance der Wieder- und Weiterverwendung deutlich.

Wichtig:

Weitere Informationen zur Pflege der Metadaten entnehmen Sie bitte dem Leitfaden Metadaten.

5. Freie Lizenzen

Werden Daten über öffentlich zugängliche Netze (das Internet) veröffentlicht, bedeutet dies noch nicht, dass diese Daten beliebig weiterverwendet werden dürfen. Urheberrechte gelten auch hier und es bedarf einer verbindlichen Regelung zur Nutzung und Verarbeitung von Daten. Die einfachste und effektivste Möglichkeit eine freie Weiterverwendung rechtssicher zu ermöglichen, ist die Festlegung von Lizenzen.

5.1. Wie lässt sich die richtige Lizenz finden?

Um den Aufwand zu minimieren empfiehlt es sich auf bestehende und vorgefertigte Lizenzmodelle zurückzugreifen.

Zwei, für Open Data geeignete, Lizenzen sind die auf den deutschen Rechtsraum angepasste Datenlizenz Deutschland 2.0 und die international bekannte Creative Commons Lizenz 4.0³.

5.2. Datenlizenz Deutschland

In einer Zusammenarbeit von Bund, Ländern und kommunalen Spitzenverbänden wurde die Datenlizenz Deutschland entwickelt. Ziel war eine einheitliche Nutzungsbestimmung für Verwaltungsdaten in Deutschland zu schaffen. Diese Lizenz ist durch den Sachverständigenrat der Open Definition als offene Lizenz anerkannt⁴ und auf die Nutzung durch Verwaltungen ausgelegt.

Die Datenlizenz Deutschland liegt aktuell in der Version 2.0 in zwei Ausfertigungen vor:

- Datenlizenz Deutschland – Namensnennung – Version 2.0 (dl-de/by-2-0-)

welche den Datennutzer verpflichtet, den jeweiligen Datenbereiter zu nennen und

³ Creative Commons (Stand 05.06.2020)

⁴ Vgl. Open Definition.org (Stand 02.06.2020).

- Datenlizenz Deutschland – Zero – Version 2.0 (dl-de/zero-2-0)

welche eine einschränkungslose Weiterverwendung ermöglicht.

Sie ist vollständig an den deutschen Rechtsraum angepasst und ermöglicht eine freie Weiterverwendung der Daten nach Open-Data-Kriterien bei bester Rechtssicherheit für Nutzer und Bereitsteller.

Die Datenlizenz Deutschland 2.0 schreibt Nutzenden nicht vor, mit welcher Lizenz ihre Neunutzung der Daten versehen sein muss.

5.3. Creative Commons Lizenzen

Creative Commons (CC) ist eine Non-Profit-Organisation, die in Form vorgefertigter Lizenzverträge eine Hilfestellung für Urheber zur Freigabe rechtlich geschützter Inhalte anbietet. Der Vorteil ist, dass die CC-Lizenzen weltweit genutzt werden. Seit Version 4.0 gibt es länderspezifische Übersetzungen der international einheitlichen CC-Lizenzen.

Insgesamt bietet CC sechs verschiedene Standard-Lizenzverträge an, die bei der Verbreitung kreativer Inhalte genutzt werden können, um die rechtlichen Bedingungen festzulegen.

Für die Verwaltung relevant, da sie die größtmögliche freie Verwendung der Daten ermöglichen sind die CC-BY und die CC-BY-SA.

Die CC-BY 4.0 bestimmt, dass bei der Verwendung der Daten der Name des Datenbereitstellers (Namensnennung) erfolgen muss. Die CC-BY-SA 4.0 verlangt die Weiterverwendung der Daten unter gleicher Lizenz.⁵

Wenn Sie an tiefer gehenden Informationen zu den Lizenzen interessiert sind, empfehlen wir das rechtliche Kurzgutachten zu Lizenzen für Open Government Data, welches das Ministerium für Wirtschaft, Innovation, Digitalisierung und Energie des Landes Nordrhein-Westfalen herausgebracht hat. Es stellt Gemeinsamkeiten und Unterschiede der Deutschlandlizenz 2.0 mit der Creative Commons Lizenz 4.0 gegenüber, zeigt deren Stärken sowie worauf geachtet werden sollte.⁶

Das BMI, GovData und das CCOD empfehlen die Nutzung der Datenlizenz Deutschland 2.0.¹

6. Wo erhalte ich Support?

BVA – Kompetenzzentrum Open Data

OpenData@bva.bund.de

<https://www.OpenData.bund.de>

SIB – Social Intranet des Bundes, Fachnetzwerk Open Data

[Verwaltung-Innovativ.de](https://www.verwaltung-innovativ.de)

⁵ Creative Commons.com (Stand 05.06.2020)

⁶ Open.NRW, Datenlizenzen für Open Government Data „Rechtliches Kurzgutachten“, 2019

7. Anhang

Liste zur Unterstützung des Datenscreenings					
Reihenfolge	Arbeitsauftrag	Verantwortlicher	Umzusetzen bis	Erledigt	Kommentar
Arbeitsschritt 1: Datenscreening					
1.	Daten nach dem 13.07.2017 erstellt?	Verfahrensanwender, Verantwortliche/-r	Open-Data-		Falls zuvor erstellt, nicht zu veröffentlichen!
2	Daten nach dem 13.07.2017 bearbeitet	Verfahrensanwender, Verantwortliche/-r	Open-Data-		Falls ja, möglicherweise zu veröffentlichen
3.	Bezeichnung des Datensatzes	Verfahrensanwender, Verantwortliche/-r	Open-Data-		
4	Wer ist der Datenverantwortliche für diese Daten?	Verfahrensanwender, Verantwortliche/-r	Open-Data-		
5	Wo werden diese Daten derzeit gespeichert?	Verfahrensanwender, Verantwortliche/-r	Open-Data-		
5.1.	In welchem Format werden sie gespeichert?	Verfahrensanwender, Verantwortliche/-r	Open-Data-		
6.	Wie werden die Daten erhoben?	Verfahrensanwender, Verantwortliche/-r	Open-Data-		
6.1.	Manuell?	Verfahrensanwender, Verantwortliche/-r	Open-Data-		

6.2.	Maschinell/ Fachverfahren?	Verfahrensanwender, Verantwortliche/-r	Open-Data-			
7.	Wer im Haus nutzt diese Daten evtl. noch?	Verfahrensanwender, Verantwortliche/-r	Open-Data-			
Reihenfolge	Arbeitsauftrag	Verantwortlicher		Umzusetzen bis	Erledigt	Kommentar
Arbeitsschritt 2: Auswahlkriterien						
8	Werden diese Daten bereits veröffentlicht?	Verfahrensanwender, Verantwortliche/-r	Open-Data-			
8.1.	Falls ja, wo und wie?	Verfahrensanwender, Verantwortliche/-r	Open-Data-			Falls ja, brauchen diese nicht veröffentlicht zu werden, sie können es aber!
9.	Besteht das uneingeschränkte Eigentum an diesen Daten?	Verfahrensanwender, Verantwortliche/-r	Open-Data-			Falls ja, können diese veröffentlicht werden.
9.2.	Falls nicht, kann die Zustimmung zur Veröffentlichung eingeholt werden?	Verfahrensanwender, Verantwortliche/-r	Open-Data-			Falls eine Zustimmung zur Veröffentlichung nicht eingeholt werden kann, nicht zu veröffentlichen!
10	Unterliegen die Daten der Geheimhaltung?	Verfahrensanwender, Verantwortliche/-r	Open-Data-			Falls ja, nicht zu veröffentlichen!
11.	Handelt es sich um personenbezogene Daten bzw. lassen sich Rückschlüsse auf Personen	Verfahrensanwender, Verantwortliche/-r	Open-Data-			Falls ja, nicht zu veröffentlichen! Ausnahme: eine Schwärzung oder Anonymi-

	daraus ableiten?				sierung ist möglich.
11.	Handelt es sich um unternehmensbezogene Daten bzw. lassen sich Rückschlüsse auf Unternehmen daraus ableiten?	Verfahrensanwender, Verantwortliche/-r	Open-Data-		Falls ja, nicht zu veröffentlichen! Ausnahme: eine Schwärzung oder Anonymisierung ist möglich.
12.	Ist der Aufwand zur Veröffentlichung überschaubar?	Verfahrensanwender, Verantwortliche/-r	Open-Data-		
13.	Wie hoch ist der Nutzen für alle Zielgruppen - geschätzt?	Verfahrensanwender, Verantwortliche/-r	Open-Data-		
14.	Wie ist die Datenqualität (Vollständigkeit, Genauigkeit, zeitlicher Bezug)?	Verfahrensanwender, Verantwortliche/-r	Open-Data-		

