



Anhang 2:

Leitfaden Anforderungen an die Daten

Version 1.4
September 2018

Inhaltsverzeichnis

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

Einleitung

Dieser Leitfaden unterstützt bei der:

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

Dieser Leitfaden beschäftigt sich mit formalen, rechtlichen und technischen Anforderungen an die bereitzustellenden Roh- und Metadaten. 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 gefunden, 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, damit sie für alle nutzbar sind.

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 empfehlendes Papier veröffentlicht. Aussagen daraus können nicht im Rahmen gesetzlicher Verpflichtungen und deren Umsetzung geltend gemacht werden.

1. Datenscreening

Zu Beginn ist es notwendig, erst einmal zu ermitteln, welche Daten seit dem 13.07.2017 erstmalig erstellt oder aktualisiert wurden.

Hierzu bietet sich entweder die manuelle Ermittlung an, wo Sie in Ihrem Haus die Fachabteilungen einzeln anfragen, welche Daten dort im eigentlichen Auftrag der Behörde erstellt bzw. aktualisiert wurden. Alternativ kann mit Hilfe eines Crawlers die Webseite Ihrer Behörde nach Daten durchsucht werden.

Anschließend empfiehlt es sich die Ergebnisse zu strukturieren. Eine Übersicht über alle in Ihrer Behörde erstellten Daten, die im eigentlichen Auftrag der Behörde erstellt wurden, kann Ihnen später nützlich sein. Finden Sie dazu in der beigefügten Checkliste eine mögliche Struktur.

Zur besseren 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.

Wählen Sie anschließend die Daten aus, die veröffentlichungsfähig sind und priorisieren Sie diese. Hierzu kann Ihnen die Übersicht von Kapitel 3.1.4.3 Auswahl über interne Priorisierung behilflich sein.

Kriterium	Erläuterung	Bewertung (Punkte 0-5)
Geheimhaltung/ rechtliche Hinder- nisse	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 Daten- qualität	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 Verfüg- barkeit	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

Sollten einzelne Kriterien mit 0 bewertet werden (rot gekennzeichnet), so sind dies Ausschließungsgründe für eine Veröffentlichung. Kriterien, bei denen 0 nicht vorkommt, können zwar kumuliert ebenso wenige Punkte erhalten, dies stellt allerdings keinen Ausschließungsgrund dar. Daten, die unter ein gelbes Kriterium fallen (gelb gekennzeichnet), müssen gemäß dem EGovG nicht veröffentlicht werden, auch wenn dies nutzbringend wäre. Diese Daten können nach entsprechender Bearbeitung ggf. veröffentlicht werden.

Falls im ersten Schritt, aufgrund einer zu geringen Bewertung, von einer Veröffentlichung abgesehen wird, müssen 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.¹

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 man kein Spezialprogramm benötigt, um die Informationen dieser Datei verarbeiten zu können. Eine Excel-Datei ist demnach zwar ein strukturiertes, jedoch kein offenes Dateiformat, da man ein Spezialprogramm benötigt, um die Datei zu öffnen. Dies schränkt den Kreis der möglichen Nutzer ein. Komma-getrennte 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:

- Sachdatenformate: 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.

¹ 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 29.12.2017).

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 Datenbereitsteller an GovData.de anzubinden, benötigen Sie als Open-Data-Bbeauftragter Ihres Ministeriums ein eigens wie ein Datenbereitstellerkonto zur Weitergabe. Ihr eigenes Konto auf GovData.de nutzen Sie für eigene Eingaben bzw. zur möglichen Korrektur von Einträgen Ihres Geschäftsbereiches. Das allgemeine Datenbereitstellerkonto dient zur Weitergabe an Mitarbeiter in Ihrem Ministerium sowie an Open-Data-Bereitsteller aus den nachgeordneten Behörden Ihres Ministeriums.

Zurzeit ist es nicht angedacht, dass jeder Mitarbeiter, der Datensätze bereit stellt, ein eigenes Konto auf GovData.de erhält. Da ein Kontozugang weitergegeben werden kann, ist hier eine entsprechende Koordination seitens des Open-Data-Bbeauftragten des Ministeriums angedacht.

Sind Sie Mitarbeiter einer nachgeordneten Behörde wenden Sie sich bitte an den Open-Data-beauftragten Ihres Ministeriums.

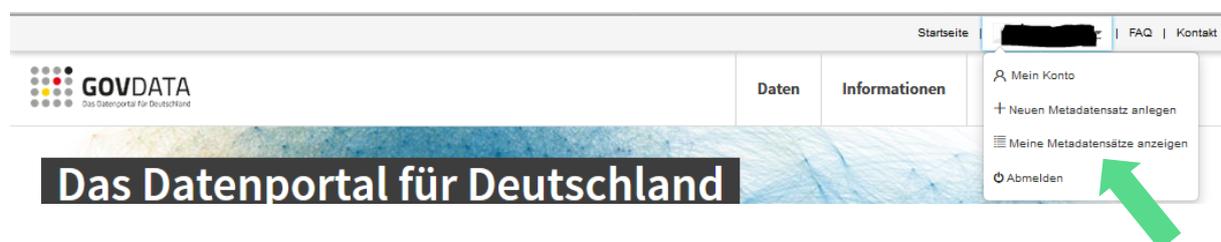
3.1. Metadaten erfassen

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

1. (Manuelle) Bereitstellung per Webformular

Dies ist die einfachste Methode Sie bietet sich an, wenn nur wenige Datensätze bereitgestellt werden sollen.

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.

Die Arbeiten für ein Datenportal laufen.

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 veröffentliche 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 neue deutsche Metadatenstandard „DCAT-AP.de“ ist mit dem europäischen Metadatenstandard „DCAT-AP“ kompatibel, wodurch eine Kompatibilität „nach Oben“ zum [Open-Data-Portal der Europäischen Union](#) gewährleistet ist.

Die Metadatenstruktur von GovData.de wurde entsprechend angepasst und vom IT-Planungsrat als verbindliche Nutzung des Interoperabilitätsstandards DCAT-AP beschlossen. Bis zum 31.12.2018 haben die an GovData.de anliefernden Bundesländer diesen Standard ebenfalls umzusetzen.

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 auffindbar zu machen. Sie enthalten die wichtigsten Informationen über den jeweiligen Datensatz sowie einen eindeutigen Verweis auf die Ressource, unter der die Daten zu finden sind.

Bei den Verweisen handelt es sich zumeist um Weblinks, wodurch eine Ressource eindeutig identifizierbar ist und die Gefahr Doubletten anzufertigen minimiert wird. 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 Datenbereitsteller 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 findet man die richtige Lizenz für seine Daten?

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 3.0 und Version 4.0 (bisher nicht portiert für Deutschland)².

5.2. Datenlizenz Deutschland

In einer Zusammenarbeit zwischen Bund, Ländern und kommunalen Spitzenverbänden wurde die Datenlizenz Deutschland entwickelt. Ziel war eine einheitliche Nutzungsbestimmungen für Verwaltungsdaten in Deutschland zu schaffen. Diese Lizenz ist durch den Sachverständigenrat der Open Definition als offene Lizenz anerkannt³ und kann somit problemlos genutzt werden.

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

- Datenlizenz Deutschland – Namensnennung – Version 2.0
welche den Datennutzer verpflichtet, den jeweiligen Datenbereinsteller zu nennen und
- Datenlizenz Deutschland – Zero – Version 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. Deshalb empfiehlt sich die Nutzung dieser Lizenz insbesondere für Behörden. Sie ist so formuliert, dass sie auch von Nicht-Juristen leicht verstanden werden kann, was die Wahrscheinlichkeit der Weiterverwendung von Daten auf Nutzerseite erhöht.⁴

Daneben gibt es eine Reihe weiterer freier Lizenzen wie beispielsweise die Creative Common Lizenz.

Das BMI empfiehlt die Nutzung der Datenlizenz Deutschland 2.0.¹

6. Wo erhalte ich Support?

BVA – Beratungszentrum des Bundes:

OpenData@bva.bund.de

² <https://de.creativecommons.org/index.php/was-ist-cc/>

³ Vgl.: GovData.de: Lizenzen, <https://www.govdata.de/web/guest/lizenzen> (Stand 26.10.2017).

⁴ GovData.de: Lizenzen. <https://www.govdata.de/web/guest/lizenzen> (Stand 19.10.2017).

SIB – Social Intranet des Bundes, Fachnetzwerk Open Data

Verwaltung-Innovativ.de

https://www.verwaltung-innovativ.de/DE/E_Government/Open_Data/open_data_node.html

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-			
Reihen- folge	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 zu veröffentlicht werden!
9.	Besteht das uneingeschränkte Eigentum an diesen Daten?	Verfahrensanwender, Verantwortliche/-r	Open-Data-			
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 daraus ableiten?	Verfahrensanwender, Verantwortliche/-r	Open-Data-			Falls ja, nicht zu veröffentlichen!

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!
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-			