



Schleswig-Holstein
Ministerium für Energiewende,
Landwirtschaft, Umwelt, Natur
und Digitalisierung

Open-Data-Leitfaden Schleswig-Holstein



Schleswig-Holstein. Der echte Norden.

Inhaltsverzeichnis

| | |
|---|-----------|
| 1 Wozu dient dieser Leitfaden? | 3 |
| 2 Welche Daten als Open Data? | 5 |
| 2.1 Was sind Daten? Welche Daten eignen sich als Open Data? | 5 |
| 2.2 Wie können Daten identifiziert werden? | 7 |
| 2.2.1 Selbst-Screening | 7 |
| 2.2.2 Fremd-Screening | 8 |
| 2.2.3 Stakeholder-Screening | 8 |
| 2.2.4 Orientierung an thematischen Datenkategorien . . | 8 |
| 2.3 Erstellung eines Datenkatalogs | 8 |
| 2.3.1 Daten, keine Dokumente | 9 |
| 2.3.2 Keine personenbezogenen Daten | 9 |
| 2.3.3 Nur bestehende Daten | 9 |
| 2.3.4 Vollständige Datensätze | 9 |
| 2.3.5 Unbearbeitete Datensätze | 9 |
| 2.3.6 Daten aus Primärquellen | 9 |
| 2.3.7 Zeitliche Nähe | 10 |
| 2.3.8 Zeitreihen | 10 |
| 3 Rechtliche Prüfung | 11 |
| 3.1 Darf ich die identifizierten Daten bereitstellen? | 11 |
| 3.2 Prüfpflichten im Detail | 12 |
| 3.2.1 Schutz entgegenstehender öffentlicher Interessen (§ 9 IZG-SH) | 12 |
| 3.2.2 Schutz personenbezogener Daten (§ 10 S. 1 Nr. 1 IZG-SH) | 12 |
| 3.2.3 Schutz geistigen Eigentums und von Urheberrechten Dritter (§ 10 S. 1 Nr. 2 u. 3. IZG-SH) | 13 |
| 4 Wie sollten Daten bereitgestellt werden? | 15 |
| 4.1 Offene Formate | 17 |
| 4.2 Datenqualität: Erläuterung 5-Sterne-Modell | 17 |
| 4.2.1 ★ Offene Lizenz (OL) | 17 |
| 4.2.2 ★★ OL + Maschinenlesbares Format (RE) | 18 |
| 4.2.3 ★★★ OL + RE + Offenes Format (OF) | 18 |
| 4.2.4 ★★★★ OL + RE + OF + Eindeutige Identifizierbarkeit (URI) | 19 |
| 4.2.5 ★★★★★ OL + RE + OF + URI + Linked Open Data (LD) | 19 |

| | |
|--|-----------|
| 4.3 Datensatz oder Datei? | 19 |
| 5 Erläuterung Metadaten | 21 |
| 5.1 Was sind Metadaten und wofür sind sie wichtig? | 21 |
| 5.2 Metadatenstandard in Deutschland: DCAT-AP.de | 22 |
| 5.3 Metadaten des Datensatzes (Dataset) | 24 |
| 5.3.1 Titel des Datensatzes | 24 |
| 5.3.2 Beschreibung | 26 |
| 5.3.3 Kategorien | 26 |
| 5.3.4 Lizenz | 27 |
| 5.3.5 Herausgeber | 27 |
| 5.3.6 Raumbezug | 28 |
| 5.3.7 Veröffentlichungszeitpunkt | 28 |
| 5.3.8 Zeitlicher Bezug | 29 |
| 5.3.9 Schlagworte | 29 |
| 5.3.10 Metadaten der Datei (Distribution) | 30 |
| 6 Lizenzen | 33 |
| 6.1 Was sind Lizenzen? | 33 |
| 6.2 Prüfpflichten vor der Lizenzauswahl | 34 |
| 6.3 Lizenzauswahl: Creative-Commons-Lizenzen | 35 |
| 6.4 Lizenz-Empfehlung für Schleswig-Holstein: CC BY 4.0 | 35 |
| 7 Bereitstellung von Daten im Open-Data-Portal Schleswig-Holstein | 37 |
| 7.1 Technische Bereitstellung von Daten | 37 |
| 7.2 Wege von Daten und Metadaten in das Portal | 37 |
| 7.2.1 Welcher Weg sollte gewählt werden? | 39 |
| 7.3 Praktische Beispiele für die Wege in das Open-Data-Portal | 40 |
| 7.3.1 Denkmalliste | 40 |
| 7.3.2 Badegewässerqualität | 40 |
| 7.4 Änderungen an veröffentlichten Datensätzen | 41 |
| 7.4.1 Korrektur von Fehlern in Datensätzen | 41 |
| 7.4.2 Reguläre Aktualisierung der Daten | 41 |
| 8 Ende des Open Data Leitfadens für Schleswig-Holstein | 43 |
| Glossar | 45 |
| Quellenverzeichnis | 49 |
| Bildverzeichnis | 51 |
| Impressum | 53 |

Kapitel 1

Wozu dient dieser Leitfaden?

Der vorliegende Leitfaden soll eine Orientierung darüber geben, was unter offenen Daten verstanden wird, wie sie zu finden sind und wie sie am Ende öffentlich zugänglich bereitgestellt werden können. Dazu enthält der Leitfaden detaillierte Schritt für Schritt-Anleitungen und Hintergrundinformationen zur Identifizierung und Bereitstellung offener Daten. Am Ende des Leitfadens finden Sie ein Glossar zu wichtigen Begriffen rund um Open Data.

Um bei der Durchsicht des Leitfadens und bei der Umsetzung von Open Data in Ihrem Arbeitsalltag eine Orientierung zu haben, finden Sie als Wegweiser vor allen Kapiteln eine Open-Data-Landkarte, die hier in der Gesamtsicht zu sehen ist. Viel Spaß beim Einstieg in das Thema Open Data!

Kapitel 2

Welche Daten als Open Data?

In diesem Kapitel erhalten Sie einen Überblick darüber, welche Daten sich grundsätzlich als Open Data eignen und wie Sie Daten identifizieren und katalogisieren können. Dazu werden Daten zunächst begrifflich von Informationen und Wissen abgegrenzt. Anschließend wird praktisch erläutert, wie Sie bei der Identifizierung von Daten vorgehen und wie ein Datenkatalog erstellt werden kann.

2.1 Was sind Daten? Welche Daten eignen sich als Open Data?

Bevor man sich auf die Suche nach Daten für eine mögliche Veröffentlichung als Open Data machen kann, sollte natürlich klar sein, wonach gesucht werden muss. Deswegen wird im Folgenden zunächst einmal erläutert, worum es sich bei Daten allgemein handelt und was aus diesen entstehen kann.

Ganz allgemein gesagt sind Daten der Rohstoff, aus dem sich Informationen entwickeln und sich Wissen generieren lässt. Daten werden zu Informationen, wenn sie in einen Kontext gestellt werden. Aus den daraus gewonnenen Informationen leitet sich wiederum Wissen ab. Der Aufbau von Wissen ist dabei ein Prozess, in dem Informationen in Entscheidungen umgewandelt werden. Die folgende Infobox verdeutlicht noch einmal die Zusammenhänge von Daten, Informationen und Wissen.

Beispiel zum Zusammenhang von Daten, Informationen und Wissen

Stellen Sie sich das Beispiel einer Urlaubsplanung vor: Daten können

hier Bewertungen, Bilder oder Preise zu Urlaubsorten sein. Die Sammlung und Präsentation solcher Daten dient der Informationsbildung. Gesammelte Informationen verdichten sich zu Wissen über Urlaubsorte und können durch einen Vergleich eine Entscheidung für oder gegen einen Urlaubsort herbeiführen.

Und was bedeutet dies nun praktisch für die alltägliche Arbeit?

Rein technisch gesehen gibt es eine Vielzahl von Daten, mit denen Sie in Ihrem Arbeitsalltag in Berührung kommen können. Hierzu zählen zum Beispiel die Internet-Seiten Ihrer Behörde oder Organisationseinheit sowie gewöhnliche PDF-Dokumente.

Bei Open Data sind aber insbesondere strukturierte Daten gemeint, also beispielsweise Daten, die in Form einer Tabelle vorliegen oder solche, die mittels eines Diagramms oder auf einer Karte visualisiert wurden. Der wichtige Kernaspekt hierbei ist, dass der Fokus bei Open Data immer auf nicht-personenbezogenen Daten liegt, d.h. auf Daten, die keine Informationen über einzelne Personen enthalten oder Rückschlüsse erlauben. Sofern Datensätze einzelne personenbezogene Daten enthalten, können diese Informationen entfernt werden, so dass der Datensatz dennoch bereinigt als Open Data veröffentlicht werden kann. Beispielsweise kann in einer Tabelle die Spalte mit personenbezogenen Daten gelöscht werden.

Um ein erstes Verständnis darüber zu erlangen, was für Daten dies konkret sein können, bietet es sich insbesondere an, bereits als Open Data veröffentlichte Beispieldatensätze zu betrachten. Diese können als Orientierung dienen. Hierzu bieten sich das Open-Data-Portal Schleswig-Holstein, sowie Open-Data-Portale anderer Bundesländer oder Kommunen, GovData.de, das Datenportal für Deutschland und der Musterdatenkatalog für Kommunen an. Im Folgenden finden Sie eine Auflistung verschiedener Open-Data-Portale:

| Bezeichnung Datenportal | Link |
|---|---|
| Open-Data-Portal Schleswig-Holstein | https://opendata.schleswig-holstein.de |
| GovData.de, Datenportal für Deutschland | https://www.govdata.de |
| Open-Data-Portal Nordrhein-Westfalen | https://open.nrw/opendata |
| Open-Data-Portal Rheinland-Pfalz | https://daten.rlp.de |
| Berliner Open-Data-Portal | https://daten.berlin.de |

Musterdatenkatalog für Kommunen

Der Musterdatenkatalog gibt Kommunen einen Anhaltspunkt, welche Daten sie als Open Data veröffentlichen könnten. Der Katalog verschafft eine gute Übersicht über Open Data und umfasst alle Kom-

munen in Deutschland, deren offene Daten über das Datenportal GovData.de auffindbar sind. Mit dem Katalog kann schnell in Erfahrung gebracht werden, welche Daten von anderen Kommunen als Open Data veröffentlicht werden. Auf diese Weise kann ein Musterdaten-katalog Initialzündung für die Öffnung von kommunalen Daten sein.

2.2 Wie können Daten identifiziert werden?

Zur Identifizierung von Datensätzen zur Veröffentlichung als Open Data kommen grundsätzlich drei unterschiedliche Formen eines „Datenscreenings“¹ (Datenrecherche) in Frage, die im Folgenden erläutert werden.

Hinweis zur Datenidentifizierung und -katalogisierung:

Es besteht kein zwingendes Erfordernis, dass Sie die in diesem und dem folgenden Kapitel erläuterten, formalen Prozesse zur Identifizierung und Katalogisierung von Datensätzen vollumfänglich befolgen. Die Identifizierung von Datensätzen kann zunächst auch niederschwellig in kleinen Schritten erfolgen. So können Sie zunächst beispielsweise auch nur einige wenige, Ihnen gut bekannte oder stark nachgefragte Datensätze erfassen und als Open Data veröffentlichen. Oberstes Ziel sollte die Identifizierung von Datensätzen, unabhängig von der Anzahl und Erfassungsmethode, sein.

2.2.1 Selbst-Screening

Bei dieser Form des Screenings geht es darum, in der eigenen Organisationseinheit oder öffentlichen Stelle zu überprüfen, welche Informationen bereits über das Internet veröffentlicht werden. Dies können alle Arten von Dokumenten, wie z. B. Artikel, Berichte oder Statistiken sein. Häufig erfüllen solche Dokumente noch nicht die Open-Data-Anforderungen, bieten aber das Potenzial, zukünftig in neuen Formaten als Open Data veröffentlicht oder um offene Daten ergänzt zu werden. Oftmals lohnt sich die nähere Begutachtung von Dokumenten, die Tabellen und Übersichten enthalten, da die verwendeten Rohdaten, also die Daten, auf denen die Tabellen und Übersichten basieren, möglicherweise als Open Data veröffentlicht werden können.

2.2.2 Fremd-Screening

Bei einem Fremd-Screening stehen andere öffentliche Stellen im Fokus; denn hier geht es um die Frage: „Welche Datensätze haben andere bereits veröffentlicht?“. Wie im vorherigen Kapitel bereits erläutert wurde, können hier verschiedene Open-Data-Portale oder der Musterdatenkatalog Orientierung bieten. Mithilfe der Orientierung an anderen Institutionen können Sie intern gezielt nach vergleichbaren Datensätzen suchen und diese für eine zukünftige Veröffentlichung erfassen.

2.2.3 Stakeholder-Screening

Vielleicht erhalten Sie in Ihrem Arbeitsalltag immer wieder Anfragen nach Daten oder Informationen von Externen. Oder sie haben sehr gute Kenntnisse über die Informationsinteressen Ihrer Kunden. Bei dieser Form des Datenscreenings geht es genau um diese Bedarfe externer Interessenten (Stakeholder). Um diesen Wissensbedarf zu erfassen, bieten sich insbesondere auch Umfragen oder Veranstaltungen an. Auch hier sollten Informationen zusammengetragen und als Hilfsmittel für die interne Datensuche verwendet werden.

2.2.4 Orientierung an thematischen Datenkategorien

Unterstützt werden können diese Datenscreenings durch eine Orientierung an den gängigen thematischen Datenkategorien, die sich in den einschlägigen Open-Data-Portalen wiederfinden lassen. Auf Grundlage dieser Kategorien kann verwaltungsintern nach Organisationseinheiten gesucht werden, die möglicherweise über Daten zu einer oder mehrerer dieser Kategorien verfügen. Die Kategorien sind im Einzelnen:

2.3 Erstellung eines Datenkatalogs

Das Ergebnis der Durchführung eines oder mehrerer Datenscreenings sollte optimalerweise, aber nicht zwingend, ein Datenkatalog sein, in dem alle zur Veröffentlichung in Frage kommenden Datensätze erfasst werden. In der einfachsten Form kann dies eine Excel-Tabelle sein, in der lediglich die identifizierten Datensätze gesammelt werden. Je nach technischen Möglichkeiten bietet sich jedoch auch die Nutzung einer MS-Access-Datenbank oder spezieller Katalogsoftware oder weiterer Fachverfahren an. Idealerweise wird ein Datenkatalog nicht nur einmalig erstellt, sondern fortlaufend aktualisiert und mit neuen Datensätzen befüllt. Folgende Informationen sollte ein Datenkatalog enthalten:

- Bezeichnung des Datensatzes

- datenliefernde Stelle (Organisationseinheit)
- datenverantwortliche Person (Kontaktdaten)
- Primärquelle des Datensatzes
- Datei-Format des gespeicherten Datensatzes
- Zeitlicher Bezug des Datensatzes
- ggf. gewünschter Veröffentlichungszeitpunkt

Bei der Erstellung eines solchen Datenkatalogs und der dafür erforderlichen Sammlung von Daten gilt es verschiedene Aspekte zu beachten, die in der nachfolgenden Übersicht dargestellt werden. Diese sind explizit nicht als Ausschlusskriterien zu verstehen, sondern lediglich als grobe Zielwerte.

2.3.1 Daten, keine Dokumente

Reine Dokumente, z. B. in Form eines PDF-Dokuments, sind keine strukturierten Daten; sie können aber solche Daten enthalten, z. B. in einer Tabelle. Hier gilt es die jeweilige Datenquelle zu identifizieren und zu veröffentlichen.

2.3.2 Keine personenbezogenen Daten

Daten, die unmittelbare Rückschlüsse auf einzelne Personen zulassen, können nicht veröffentlicht werden.

2.3.3 Nur bestehende Daten

Es sollten Daten veröffentlicht werden, die ohnehin bereits vorliegen und nicht extra erhoben werden müssen.

2.3.4 Vollständige Datensätze

Zu veröffentlichende Datensätze sollten so vollständig wie möglich sein, um den Umfang eines Themas auch möglichst umfassend abbilden zu können.

2.3.5 Unbearbeitete Datensätze

Daten sollten mit größtmöglichem Feinheitsgrad gesammelt und veröffentlicht werden.

2.3.6 Daten aus Primärquellen

Daten sind an ihrem Ursprungsort zu identifizieren und zu sammeln.

2.3.7 Zeitliche Nähe

Grundsätzlich sollten Daten möglichst zeitnah nach Erhebung oder direkt zeitgleich zur Erhebung veröffentlicht werden, um ihre Aktualität zu gewährleisten.

2.3.8 Zeitreihen

In bestimmten Kontexten kann es auch sinnvoll sein, historische Daten für Zeitreihen zu veröffentlichen:

1. Für die Reproduzierbarkeit ist es erforderlich, auf die einer Analyse, Studie etc. zugrundeliegenden Daten zugreifen zu können. Um mögliche Veränderungen bei der Übertragung ausschließen zu können, sollten diese Daten von der Primärquelle abrufbar sein.
2. Um historische Forschung durchführen zu können, sind ältere Datenbestände unverzichtbar. Beispielsweise könnte es von Interesse sein, in welchem Jahrzehnt welche Gebäude als denkmalwürdig angesehen wurden und wie dies begründet wurde.
3. Für das Training von Künstlicher Intelligenz sind Zeitreihen erforderlich. Nur so ist es z. B. möglich, aufgrund historischer Pegelstände, Wetterdaten und ggf. weiterer Parameter ein Modell zu trainieren, das Überflutungen vorhersagen kann.

Kapitel 3

Rechtliche Prüfung

In diesem Kapitel erfahren Sie, welche rechtlichen Prüfpflichten es vor der Veröffentlichung von Daten als Open Data zu beachten gilt und insbesondere welche Schutzrechte zu wahren sind. Nach einer kurzen Einführung finden Sie im zweiten Unterkapitel dazu eine Schritt-für-Schritt-Anleitung zur Durchführung der erforderlichen rechtlichen Prüfschritte.

3.1 Darf ich die identifizierten Daten bereitstellen?

Nachdem im vorherigen Kapitel gezeigt wurde, wie geeignete Daten identifiziert werden können, erfolgt in diesem Kapitel eine Erläuterung, welche rechtlichen Aspekte es bei einer Veröffentlichung zu beachten gilt.

Bevor identifizierte Daten als Open Data veröffentlicht werden dürfen, sind verschiedene Prüfpflichten, die sich aus dem Informationszugangsgesetz für das Land Schleswig-Holstein (IZG-SH) ergeben, zu beachten. Die einzelnen Schritte und dahinterstehenden Pflichten werden im weiteren Verlauf näher dargelegt. Der nachfolgende Entscheidungsbaum vermittelt einen ersten Überblick über die existierenden Prüfpflichten, bei denen es jeweils um den Schutz unterschiedlicher Interessen und Güter geht.

Mit „Maßnahmen zur Veröffentlichung sind möglich“ ist hier gemeint, dass die Daten durch eine Bearbeitung so verändert werden können, dass keine Schutzinteressen nach dem IZG-SH mehr einer Veröffentlichung entgegenstehen. Die Bearbeitung kann in Form einer Anonymisierung, statistischen Verfremdung oder Aggregation von Daten erfolgen.

3.2 Prüfpflichten im Detail

3.2.1 Schutz entgegenstehender öffentlicher Interessen (§ 9 IZG-SH)

In einem ersten Schritt ist zu prüfen, ob öffentliche Interessen einer Veröffentlichung von Daten als Open Data entgegenstehen. In Anlehnung an § 9 Abs. 1 Nr. 1-5 IZG-SH dürfen Daten nicht veröffentlicht werden, wenn dadurch

- die internationalen Beziehungen Schleswig-Holsteins, die Verteidigung oder bedeutsame Schutzgüter der öffentlichen Sicherheit,
- die Beziehungen Schleswig-Holsteins zum Bund oder einem anderen Land,
- die Vertraulichkeit der Beratungen von informationspflichtigen Stellen,
- die Durchführung eines laufenden Gerichtsverfahrens, den Anspruch einer Person auf ein faires Verfahren oder die Durchführung strafrechtlicher, ordnungswidrigkeitenrechtlicher oder disziplinarrechtlicher Ermittlungen oder
- der Zustand der Umwelt und ihrer Bestandteile im Sinne des § 2 Abs. 2 Nr. 1 IZG-SH oder Schutzgüter im Sinne des § 2 Abs. 2 Nr. 6 IZG-SH

gefährdet werden.

Es hat im Einzelfall eine Abwägung zwischen dem öffentlichen Interesse an der Geheimhaltung und dem öffentlichen Bekanntgabeinteresse zu erfolgen.

Die aufgelisteten Aspekte sind nicht als abschließende Ausschlusskriterien zu verstehen: mitunter besteht die Möglichkeit, Daten durch eine Bearbeitung so zu verändern, dass öffentliche Interessen nicht mehr gefährdet werden. Die Bearbeitung kann in Form einer Anonymisierung, statistischen Verfremdung oder Aggregation von Daten erfolgen. Ist eine Bearbeitung erfolgt, muss dies später im Bereitstellungsprozess bei der Erfassung der Metadaten dokumentiert werden. Nach einer etwaigen Bearbeitung kann die Prüfung vorgesetzt werden.

3.2.2 Schutz personenbezogener Daten (§ 10 S. 1 Nr. 1 IZG-SH)

In einem zweiten Schritt ist zu überprüfen, ob die für eine Veröffentlichung vorgesehenen Daten einen Personenbezug aufweisen oder dieser hergestellt werden kann.

Begriffsbestimmung „personenbezogene Daten“

Personenbezogene Daten sind alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen; als iden-

tifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen identifiziert werden kann, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind (Art. 4 Nr. 1 DSGVO).

Personenbezogene Daten dürfen grundsätzlich nur dann verarbeitet (hierzu zählt auch eine Veröffentlichung) werden, wenn eine Rechtsvorschrift dies erlaubt oder die betroffene Person einer Verarbeitung eingewilligt hat (vgl. Art. 6 DSGVO). Bei der Veröffentlichung von Informationen auf Grundlage der Einwilligung einer betroffenen Person sind zudem die Kriterien an eine gültige Einwilligung (vgl. Art. 7 DSGVO) zu erfüllen. Insbesondere muss in einem Prozess die Möglichkeit der betroffenen Person die gegebene Einwilligung jederzeit für die Zukunft zu widerrufen, umgesetzt werden können.

Soweit eine Information keiner Person zugeordnet ist oder (nicht mehr) zugeordnet werden kann, liegt kein Personenbezug vor und eine Veröffentlichung ist möglich. Anonyme oder anonymisierte Daten unterfallen nicht dem Schutz der DSGVO und können damit auch veröffentlicht werden.

Auch hier bietet sich die Bearbeitung von Daten an, beispielsweise in Form des Entfernens einer Spalte mit personenbezogenen Daten aus einer Tabelle, um eine Anonymisierung zu erreichen und die Daten damit veröffentlichen zu können. Nehmen Sie bei weiteren Fragen bitte Kontakt zu Ihrer*em Datenschutzbeauftragten auf, um im Detail zu klären, welche (technischen) Möglichkeiten der Anonymisierung in Ihrer Behörde oder Organisationseinheit bestehen.

3.2.3 Schutz geistigen Eigentums und von Urheberrechten Dritter (§ 10 S. 1 Nr. 2 u. 3. IZG-SH)

Schließlich ist in einem dritten Schritt die Wahrung des Schutzes geistigen Eigentums sowie der Schutz von Urheberrechten Dritter zu prüfen. In Anlehnung an § 10 S. 1 Nr. 2 und 3 IZG-SH dürfen Daten nicht veröffentlicht werden, wenn ihre Veröffentlichung

- Rechte am geistigen Eigentum, insbesondere Urheberrechte, verletzen würden,
- Betriebs- oder Geschäftsgeheimnisse zugänglich gemacht würden oder die Informationen dem Steuer- oder Statistikgeheimnis unterliegen [...].

Hinweis: Lizenzen

Unter den Bereich der rechtlichen Prüfung fällt ebenfalls das Thema der Lizenzen, unter denen ein Datensatz genutzt werden darf. Welche Lizenzen es gibt und welche im Zusammenhang von Open Data genutzt verwendet werden sollten, wird in Kapitel 5 näher erläutert.

Kapitel 4

Wie sollten Daten bereitgestellt werden?

In diesem Kapitel erfolgt eine Erläuterung, in welcher Form Daten nach einer rechtlichen Prüfung bereitgestellt werden sollten.

Das übergeordnete Ziel bei der Bereitstellung von offenen Daten sollte die größtmögliche und gleichzeitig auch einfachste Möglichkeit der Weiterverwendung der Daten sein. Dies gilt sowohl für Menschen als auch Maschinen. Um die Weiterverwendung optimal zu gewährleisten, ist es erforderlich, dass Daten maschinenlesbar und in offenen Formaten bereitgestellt werden. Dieses Kapitel soll Ihnen einen ersten Überblick über diese beiden Aspekte geben und ein grundlegendes Verständnis vermitteln.

Am Ende dieses Kapitels erfolgt darüber hinaus eine Erklärung, welche Möglichkeiten der separierten oder gesammelten Veröffentlichung von Daten im Open-Data-Portal Schleswig-Holstein bestehen. ## Maschinenlesbarkeit

Die Maschinenlesbarkeit eines Datensatzes sagt aus, ob dieser von einer Maschine gelesen bzw. interpretiert und weiterverarbeitet werden kann. Durch die Maschinenlesbarkeit wird die einfache Einbindung von Daten in Softwareanwendungen ermöglicht und damit eine größtmögliche Nutzbarkeit der Daten erzielt. So können offene, maschinenlesbare Daten beispielsweise für Applikationen wie die Wheelmap genutzt und damit für eine Vielzahl von Personen zugänglich gemacht werden.

Beispiele für maschinenlesbare Dateiformate:

| Grad der Maschinenlesbarkeit | Formate |
|---|---|
| keine bis geringe überwiegende vollständige | PDF, DOC, DOCX, GIF, JPG, JPEG, PNG, TIFF, GeoTIFF, ODT TXT, RTF, ODS, XLS, CSV, HTML, XLSX XML, RDF, RSS, KMZ, DXF, GPX, GML, JSON |

Aus älteren Fachanwendungen können moderne Formate wie XML, RDF oder JSON teilweise nicht exportiert werden, weshalb sich das Format CSV als anzustrebender Mindeststandard für offene Daten etabliert hat.

Erläuterung: CSV

Das Format CSV steht im englischen Original für „Comma Separated Values“. Hierunter werden tabellarische Daten verstanden, die in einem Textformat gespeichert und über einen Zeichentrenner (meistens Komma oder Semikolon) getrennt hintereinandergeschrieben werden.

Tabellenkalkulationsprogramme wie Microsoft Excel oder LibreOffice Calc können Formate wie CSV abspeichern.¹ Die nachfolgende Abbildung (geöffnet in einem Text-Editor; Datensatz aus dem Open-Data-Portal Schleswig-Holstein) visualisiert den Aufbau einer CSV-Datei. Die aufgelisteten Werte werden durch Kommata getrennt; die oberste Zeile enthält die jeweiligen Spaltenüberschriften.

```
1 "Datum", "Gemeldete Fälle", "Hospitalisierungen", "Verstorben"
  "20.05.2021", 106, 2, 0
3  "19.05.2021", 172, 7, 0
  "18.05.2021", 202, 6, 0
5  "17.05.2021", 93, 4, 0
  "16.05.2021", 59, 5, 0
7  "15.05.2021", 149, 7, 0
```

Hinweis: Visuelle Optimierung von Datensätzen

Häufig werden tabellarische Daten für menschliche Betrachterinnen oder Betrachter visuell durch zusätzliche Überschriften oder Leerzeichen optimiert, was jedoch zu Schwierigkeiten und Fehlern bei der automatisierten Weiterverwendung durch Maschinen führen kann. Hier stehen also die Interpretierung von Daten durch Menschen und Maschinen im Konflikt zueinander. Im Sinne des Gebots der Maschinenlesbarkeit und der so gegebenen Auswertungsmöglichkeit, sollte jedoch dringend von einer Optimierung für menschliche Betrachterinnen und Betrachter abgesehen werden.

In dem [NQDM – Leitfaden für qualitativ hochwertige Daten und Metadaten](#) finden Sie ergänzend zu diesem Kapitel eine ausführliche Erläuterung des CSV-Formats sowie Empfehlungen zum Umgang damit. Der Leitfaden enthält auch weitere Erläuterungen und Empfehlungen zu einigen der in der Tabelle oben aufgeführten Dateiformate.

4.1 Offene Formate

Um sicherzustellen, dass jeder Person der Zugriff auf die offen bereitgestellten Daten möglich ist, sollten bei der Bereitstellung stets offene Formate verwendet werden.

Ein Dateiformat ist offen, wenn kein Spezialprogramm für die Nutzung benötigt wird und bei der Nutzung keine Kosten entstehen. Das Primärziel von Formatoffenheit ist es, den ungehinderten Umgang mit Daten zu ermöglichen, ohne rechtliche oder technische Restriktionen oder Abhängigkeiten von einem bestimmten Anbieter oder Hersteller zu kreieren.

Beispiele für offene Dateiformate⁴:

- proprietäre Formate (diese werden von privaten Interessen, z. B. einzelnen Herstellern wie Microsoft, definiert und kontrolliert): XLS, DOC, DOCX, PPT, TIFF, GeoTIFF
- nicht proprietäre, offene Formate: TXT, CSV, HTML, XML, RDF, ODT, ODS, RSS, XLSX, PDF, PNG, KMZ, GML, JSON, RTF, GIF, JPG/JPEG, DXF, GPX

Die Veröffentlichung von Daten in proprietären Formaten sollte die Ausnahme darstellen, da bei diesen Dateiformaten oben genannte Abhängigkeiten bestehen. Das Format eines Datensatzes sollte möglichst in Form eines offen und unentgeltlich nutzbaren Standards präzise definiert und dokumentiert sein. Das Vorhandensein eines offenen Standards garantiert, dass Daten in diesem Format jederzeit und von jeder Person korrekt verarbeitet werden können.¹

Oftmals liegen Daten in verschiedenen Formaten vor; um einer möglichst großen Zielgruppe von Nutzer*innen mit unterschiedlicher technischer Ausstattung die Nutzung von Daten zu ermöglichen, sollten die Daten auch in verschiedenen Formaten gleichzeitig bereitgestellt werden. Dabei ist die inhaltliche Übereinstimmung der Daten zu wahren.

4.2 Datenqualität: Erläuterung 5-Sterne-Modell

Zur Beurteilung der Qualität von Open Data kann das 5-Sterne-Modell für offene Daten [[5starinfo](#)] herangezogen werden, das nachfolgend erläutert wird. Dieses Modell wird auch in dem Open-Data-Portal Schleswig-Holstein zur Bewertung der Qualität der bereitgestellten Daten verwendet.

4.2.1 ★ Offene Lizenz (OL)

Den ersten Stern erhalten Daten, die unter einer offenen Lizenz (mit der festgelegt wird, unter welchen Bedingungen bzw. mit welchen

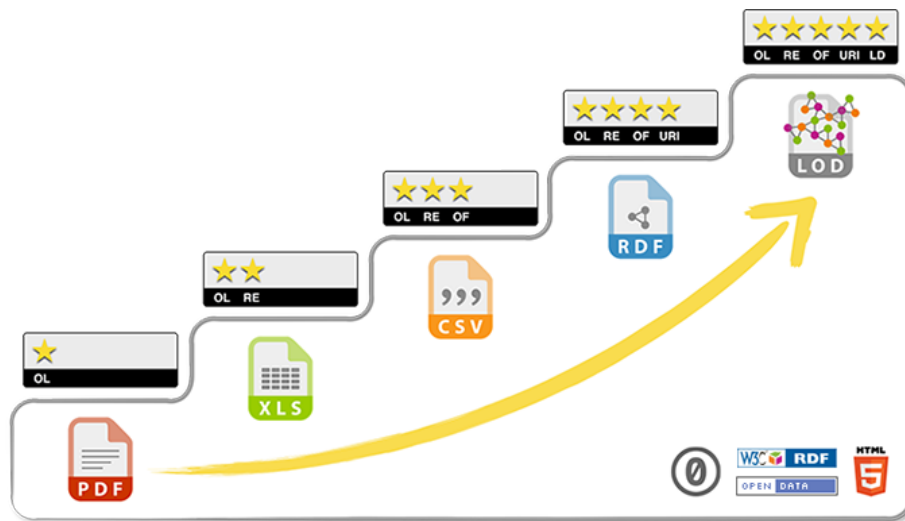


Abbildung 4.1: Stufen des 5-Sterne-Modells für offene Daten

Einschränkungen ein Datensatz verwendet werden darf) zur Verfügung gestellt werden. Das Dateiformat ist hier noch nicht von Relevanz.

Vorteile:

- Nutzende können Daten ohne Einschränkungen ansehen, lokal speichern, verändern oder teilen
- Herausgebende müssen nicht wiederholt erklären, dass bereitgestellte Daten weiterverwendet werden dürfen

4.2.2 ★★ OL + Maschinenlesbares Format (RE)

Daten, die in einem wiederverwendbaren (strukturierten) Format vorliegen erhalten einen zweiten Stern. Wiederverwendbarkeit wird durch die Verwendung von Formaten, wie z. B. Microsoft Excel gewährleistet. Beispielsweise sollte anstatt eines eingescannten Bildes eine Tabelle verwendet werden. Die Wiederverwendbarkeit ist als ein Qualitätsmerkmal von Daten zu verstehen, die auch außerhalb ihres originären Kontextes verwendet werden können.

Vorteile:

- Nutzende können Daten mit (proprietärer) Software (z. B. Microsoft Excel) bearbeiten
- das Dateiformat kann in ein anderes Format exportiert werden

4.2.3 ★★★ OL + RE + Offenes Format (OF)

Mit einem dritten Stern werden Daten ausgezeichnet, die unter einer offenen Lizenz in einem wiederverwendbaren, nicht proprietären

Format bereitgestellt werden. Dies bedeutet, dass Daten ohne proprietäre Software verwendet werden können. Beispiel: CSV-Format anstatt Excel-Format.

Vorteile:

- es muss keine (kostenpflichtige) proprietäre Software zur Nutzung der Daten verwendet werden

4.2.4 ** OL + RE + OF + Eindeutige Identifizierbarkeit (URI)**

Daten, die zusätzlich zu den vorherigen Kriterien mit einer URI (Uniform Resource Identifier) zur eindeutigen Identifizierung versehen sind, erhalten einen vierten Stern.

Vorteile:

- Nutzende können Daten lokal oder online verlinken und Lesezeichen verwenden
- Daten können mit anderen Daten kombiniert werden
- Herausgebende können mit der Verwendung von URIs den Zugang zu ihren Daten optimieren

4.2.5 *** OL + RE + OF + URI + Linked Open Data (LD)**

Alle fünf Sterne erhalten Daten, wenn sie alle vorherigen Kriterien erfüllen und darüber hinaus mit anderen Daten verknüpft sind (Linked Open Data). Durch diese Verknüpfung werden Daten in einen Kontext gestellt und eine Navigation zwischen verschiedenen Datenpunkten wird möglich.

Vorteile:

- Nutzende können während der Nutzung direkt weitere Daten entdecken und etwas über das Datenschema lernen
- Herausgebende profitieren von einer verbesserten Auffindbarkeit und Wertsteigerung ihrer Daten

4.3 Datensatz oder Datei?

Bei der Veröffentlichung von Daten stellt sich oftmals die Frage, ob einzelne zusammenhängende Daten entweder als separate Datensätze oder aber als separate Dateien eines einzelnen Datensatzes veröffentlicht werden sollten. Relevant kann dies beispielsweise bei Zeitreihen (gleichartige Daten, die zu unterschiedlichen Zeitpunkten erhoben wurden oder sich über einen Zeitraum erstrecken) oder bei Daten, die sich auf unterschiedliche räumliche Positionen beziehen, sein. Ein pauschale Antwort, ob Daten als einzelne Datensätze oder aber als

ein gesammelter Datensatz veröffentlicht werden sollten, gibt es nicht. Vielmehr sind unterschiedliche Formen der Veröffentlichung denkbar, die im Folgenden anhand eines Beispiels anschaulich gemacht werden sollen.

Für die Erläuterung wird auf das Beispiel der Parkplätze in Scharbeutz zurückgegriffen. Diese Datensätze haben sowohl zeitliche (jeweils ein Tag) als auch räumliche Bezüge (jeweils ein Parkplatz in Scharbeutz). Sie werden von der Tourismus-Agentur Lübecker Bucht zentral erhoben und täglich veröffentlicht. Die Rohdaten sind hierbei jeweils eine CSV-Datei mit Messwerten pro Parkplatz und Tag. Für diese Daten bestehen grundsätzlich mehrere Veröffentlichungsoptionen, die in der folgenden Tabelle dargestellt werden:

| Option | Datensatz | Im Datensatz enthaltene Dateien |
|--------|---|--|
| 1 | Parkplätze Scharbeutz | Parkplatz Scharbeutz Haffkrug P1 31.05.2021, Parkplatz Scharbeutz Badeweg 31.05.2021, etc. |
| 2 | Parkplätze Scharbeutz 31.05.2021 | Parkplatz Scharbeutz Haffkrug P1, Parkplatz Scharbeutz Badeweg, etc. |
| 3 | Parkplatz Scharbeutz Badeweg | Parkplatz Scharbeutz Badeweg 31.05.2021, Parkplatz Scharbeutz Badeweg 30.05.2021, etc. |
| 4 | Parkplatz Scharbeutz Badeweg 31.05.2021 | Hier ist die Datei identisch mit dem Datensatz. |

Grundsätzlich wären hier alle vier Veröffentlichungsoptionen möglich. Die Tourismus-Agentur Lübecker Bucht hat sich für die vierte Option entschieden. So werden täglich aktualisierte Datensätze (z. B. [Parkplatz Scharbeutz Badeweg](#)) für unterschiedliche Parkplätze in Scharbeutz veröffentlicht, aus denen anhand von Messwerten hervorgeht, zu welchem Zeitpunkt wie viele Fahrzeuge auf den Parkplätzen stehen. Die Messwerte werden in Form des „[Beachtickers](#)“ visualisiert, der von Besucherinnen und Besuchern von Scharbeutz genutzt werden kann, um zu planen, wo sie ihre Fahrzeuge parken.

Kapitel 5

Erläuterung Metadaten

In diesem Kapitel erfolgt eine Erläuterung, wie sichergestellt werden kann, dass bereitgestellte Daten für Interessierte auch auffindbar gemacht werden können. Hier sind qualitativ hochwertige Metadaten (Daten, die die eigentlichen Daten beschreiben) von größter Bedeutung. Das Kapitel enthält Handlungsempfehlungen, Hinweise und Beispiele, die als Hilfestellungen und Orientierung für die Veröffentlichungsprozesse in der täglichen Arbeit herangezogen werden können.

5.1 Was sind Metadaten und wofür sind sie wichtig?

Wie bereits in vorangegangenen Kapiteln erwähnt, handelt es sich bei Metadaten um „Daten über Daten“. Metadaten beschreiben die eigentlichen, inhaltlichen Daten.

Mithilfe von Metadaten können veröffentlichte Daten also überhaupt erst gefunden werden, was Metadaten außerordentlich wichtig macht. Denn Daten, die nicht auffindbar sind, können durch eine Weiterverwendung auch nicht ihren vollen Mehrwert entwickeln. Metadaten sind also ein wesentlicher Baustein und Grundsatz der Open-Data-Idee. Entsprechend groß sollte daher die Sorgfalt bei der Erstellung und Publikation von Metadaten sein.

In der folgenden Übersicht sehen Sie den Beispieldatensatz „Nachrichtenbroker“ mit den dazugehörigen Metadaten, die dafür sorgen, dass dieser Datensatz im Open-Data-Portal Schleswig-Holstein gefunden werden kann.

| Metadatum | Wert |
|-----------------------|-----------------------------|
| Titel des Datensatzes | Statistik Nachrichtenbroker |

| Metadatum | Wert |
|-----------------------------|---|
| Beschreibung | Die Statistik gibt Auskunft über die Anzahl der über den Nachrichtenbroker vermittelten Nachrichten je Dienstpaket. Der Nachrichtenbroker ist eine zentrale Vermittlungsstelle, die man sich als eine „Datendrehscheibe“ vorstellen kann. |
| Kategorien | Regierung und öffentlicher Sektor, Wissenschaft und Technologie |
| Lizenz | Creative Commons CC Zero License (cc-zero) |
| Herausgeber | Zentrales IT-Management Schleswig-Holstein |
| Raumbezug | Bundesland Schleswig-Holstein |
| Veröffentlichungszeitpunkt | 09.04.2021 (zuletzt bearbeitet am 09.04.2021) |
| Zeitraum (zeitlicher Bezug) | 01. Quartal 2021 |
| 5-Sterne-Bewertung | *** |
| Schlagworte | Nachrichtenbroker, XÖV |
| Link zum Datensatz | https://opendata.schleswig-holstein.de/dataset/statistik-nachrichtenbroker-2021-q1 |

Bei der geplanten Veröffentlichung von Daten ist also in organisatorischer Hinsicht nicht nur die Veröffentlichung an sich, sondern auch die Erfassung und Veröffentlichung von Metadaten durchzuführen. Zur Beschreibung von Metadaten wurde für Deutschland durch den IT-Planungsrat der Metadatenstandard „DCAT-AP.de“ verbindlich festgelegt¹. Durch die Nutzung von Metadatenstandards und einer einheitlichen Vorgehensweise, wie DCAT-AP.de, sollen Datensätze institutionsübergreifend auffindbar und durchsuchbar gemacht werden. Der Metadatenstandard wird im nachfolgenden Unterkapitel näher erläutert.

5.2 Metadatenstandard in Deutschland: DCAT-AP.de

Das „Data Catalogue Application Profile“ (DCAT-AP) für Datenportale in Europa ist eine Spezifikation, die auf dem Datenkatalog-Vokabular des W3C (World Wide Web Consortium) zur Beschreibung von Datensätzen des öffentlichen Sektors in Europa basiert. DCAT-AP.de ist die deutsche Adaption von DCAT-AP und wird als gemeinsames deut-

sches Metadatenmodell zum Austausch von offenen Verwaltungsdaten verwendet:

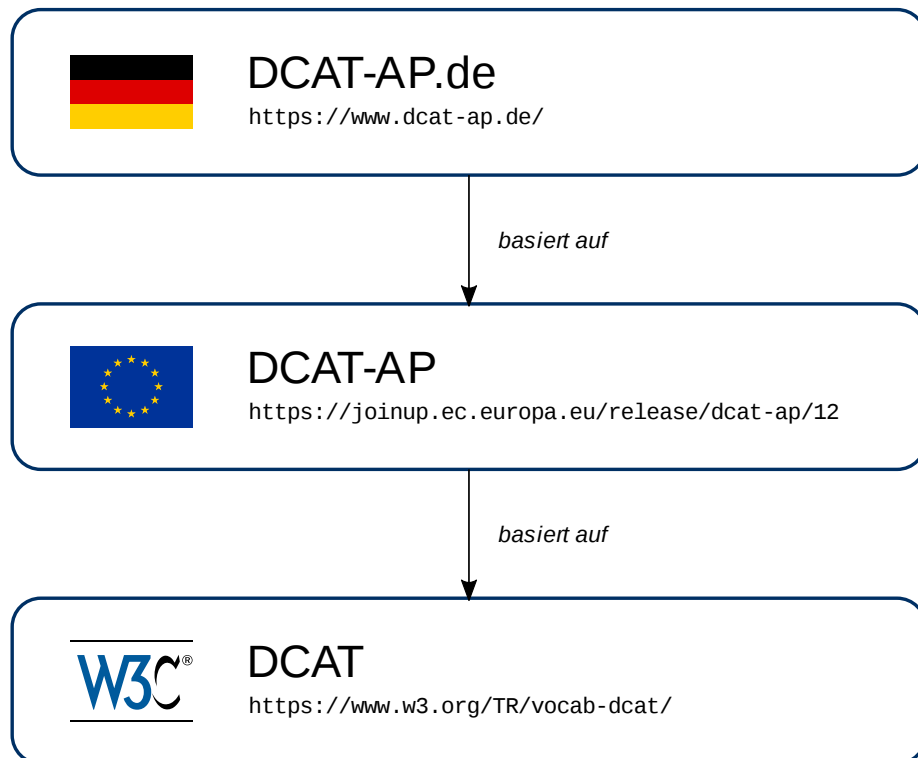


Abbildung 5.1: Das deutsche DCAT-AP.de basiert auf dem europäischen DCAT-AP, das wiederum auf dem internationalen DCAT basiert

Die beiden zentralen Elemente in DCAT-AP.de sind Dataset und Distribution.

Man kann sagen, dass ein Dataset (Datensatz) die Informationen logisch beschreibt, während eine Distribution die technische Beschreibung einer konkreten Datei ist. Hat man nur eine einzige Datei (z. B. eine CSV-Datei), so besteht das Dataset nur aus einer Distribution. Es ist jedoch auch möglich, Daten in verschiedenen Formaten (oder als Dienst + Download) anzubieten. In dem Fall hat ein Dataset mehrere Distributionen. Wichtig ist, dass sich die Distribution nur im Format unterscheiden. Unterscheidet sich auch der Inhalt, so sind mehrere Datasets anzulegen.

Erläuterung und Beispiele zum Zusammenhang von Dataset (Datensatz) und Distribution:

Bei anderen Open-Data-Portalen hat sich gezeigt, dass die Zuordnung von Datensatz und Distribution nicht immer einfach ist. Wichtig ist zunächst zu verstehen, dass ein Datensatz zwar mehrere Distributionen enthalten kann, diese dann aber den gleichen Inhalt haben müssen. Im Europäischen Datenportal wird dies sinngemäß wie folgt formuliert:

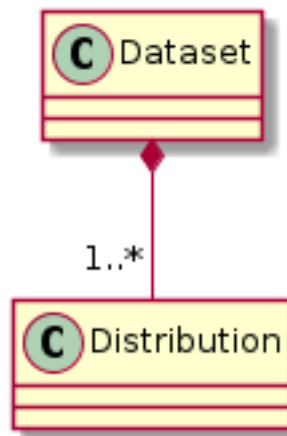


Abbildung 5.2: UML-Klassendiagramm, das den Zusammenhang zwischen Dataset und Distribution zeigt

„Wenn ein Datensatz mehr als eine Distribution (z.B. als PDF und als CSV) hat, sind alle Distributionen inhaltlich identisch. Sie unterscheiden sich nur in der Darstellung der Daten. Zum Beispiel kann ein Datensatz zwei Distributoren haben, von denen einer die Daten als PDF und einer die Daten als maschinenlesbare XML/RDF-Datei oder eben als CSV anbietet.“¹

Im Folgenden sehen Sie mehrere Beispiele welche sowohl das RICHTIGE als auch das FALSCHGE Vorgehen darstellen:

| RICHTIG | |
|---------------------|--|
| Datensatz | Statusliste Schule |
| Zeitraum | April 2021 |
| Distribution | Statusliste-April.csv + Statusliste-April.json + Statusliste-April.pdf |
| Erläuterung | Die Distributionen enthalten hier alle die gleichen Informationen, sie wurden nur in unterschiedlichen Formaten aufbereitet. |

Abbildung 5.3: Beispiel für einen richtigen Datensatz

5.3 Metadaten des Datensatzes (Dataset)

5.3.1 Titel des Datensatzes

Der Titel ist eine kurze aussagekräftige Bezeichnung des Datensatzes. Der Titel sollte grundsätzlich Angaben zum Raumbezug sowie zum zeitlichen Bezug enthalten, damit ein Datensatz bei der Auflistung

| FALSCH | |
|----------------------|---|
| Daten satz | Statusliste Schule |
| Zeitraum | ab April 2021 |
| Distrib ution | Statusliste-April.csv + Statusliste-Mai.csv + Statusliste-Juni.csv |
| Erläuterung | Die Distributionen haben hier einen unterschiedlichen Inhalt, die Liste von April hat nicht den gleichen Inhalt wie die von Mai und Juni. Richtig wäre es, drei Datensätze zu erzeugen. |

Abbildung 5.4: Beispiel für einen falsch konstruierten Datensatz

| FALSCH | |
|----------------------|--|
| Daten satz | Daten des Bauhofs |
| Zeitraum | ab Oktober 2021 |
| Distrib ution | baum kataster.csv + straß enlaternen.csv + spielplätze.csv |
| Erläuterung | Die Distributionen haben hier einen unterschiedlichen Inhalt. Richtig wäre es, vier Datensätze mit jeweils einer Distribution zu erzeugen. |

Abbildung 5.5: Beispiel für einen falsch konstruierten Datensatz

| FALSCH | |
|----------------------|--|
| Daten satz | Daten des Bauhofs |
| Zeitraum | ab Oktober 2021 |
| Distrib ution | bauhof.xlsx |
| Erläuterung | Es handelt sich hier im Prinzip nur um eine andere Version des vorherigen Problems. Auch wenn dieses Vorgehen formal DCAT-AP.de konform ist, ist es nicht zielführend, da es die Daten „versteckt“. Richtig wäre es, vier Datensätze mit jeweils einer (CSV-Datei) oder zwei (CSV-Datei + Excel-Datei) Distributionen anzulegen. |

Abbildung 5.6: Beispiel für einen falsch konstruierten Datensatz

direkt von anderen abgegrenzt werden kann. Beispiel: Tätige Personen im Verarbeitenden Gewerbe in Ostholstein (die Informationen zum zeitlichen Bezug, hier 2008-2020, sind auf der Übersichtsseite der Datensätze ersichtlich).

5.3.2 Beschreibung

Die Beschreibung fasst die Inhalte eines Datensatzes kurz zusammen. Neben dem Titel ist die Beschreibung elementar für die Auffindbarkeit und Nutzung eines Datensatzes. Nutzende können sich damit einen schnellen Überblick über bereitgestellte Daten verschaffen.

Empfehlung: Beschreibung eines Datensatzes

Die Beschreibung sollte der Nutzerin oder dem Nutzer bereits Aufschlüsse darüber geben, was in einem Datensatz zu finden ist, ohne den Datensatz geöffnet zu haben.

Versuchen Sie sich bei der Beschreibung eines Datensatzes in die Rolle einer potenziellen Nutzerin oder eines Nutzers zu versetzen. Welche Informationen benötigen diese Personen, wenn sie erstmalig auf einen Datensatz stoßen und zur Nutzung angeregt werden sollen?

Bei der Formulierung sollte darauf geachtet werden, dass Fremdwörter, „Amtsdeutsch“ oder komplexe Formulierungen die Ausnahme darstellen.

Die Beschreibung sollte u.a. die folgenden Fragen beantworten:

- Um welchen Datensatz handelt es sich?
- Welche Informationen beinhaltet der Datensatz?
- Durch wen werden die Daten erhoben und bereitgestellt?
- Welche Erhebung liegt dem Datensatz zugrunde?
- Wofür eignet sich der Datensatz **nicht**?

5.3.3 Kategorien

Zur Vereinheitlichung der Kategorisierung von Daten beinhaltet die o.g. DCAT-AP.de-Spezifikation 13 europaweit einheitliche, thematische Kategorien, unter denen Daten bei ihrer Veröffentlichung eingeordnet werden. Damit soll die Auffindbarkeit und Durchsuchbarkeit von Datensätzen erhöht werden. Diese Kategorien werden auch im Open-Data-Portal Schleswig-Holstein verwendet und sind bei einer Veröffentlichung von Daten zu berücksichtigen.

Empfehlung: Auswahl von Kategorien

Es kommt durchaus vor, dass Datensätze inhaltlich mehreren Kategorien zugeordnet werden können; dies stellt grundsätzlich auch kein Problem dar, jedoch sollten Sie bei der Auswahl der Kategorien so präzise und sparsam wie möglich vorgehen und nur die nötigsten Kategorien auswählen.

5.3.4 Lizenz

Lizenzen geben an, unter welchen Bedingungen bzw. mit welchen Einschränkungen ein Datensatz verwendet werden darf. Für Schleswig-Holstein werden durch das Zentrale IT-Management Schleswig-Holstein die Lizenzen Creative Commons Namensnennung – 4.0 International (CC BY 4.0) sowie in Ausnahmefällen unter besonderen Bedingungen Creative Commons Namensnennung - Nicht kommerziell 4.0 International (CC BY-NC 4.0) empfohlen. Näheres zu dem Thema Lizenzen und wie diese ausgewählt werden und welche rechtlichen Prüfschritte es vor der Lizenzauswahl gibt, erfahren Sie in Kapitel 6 dieses Leitfadens.

5.3.5 Herausgeber

Um nachvollziehen zu können, von wem Daten veröffentlicht wurden und an wen sich Datennutzende bei Fragen wenden können, muss der Herausgeber der Daten bei der Veröffentlichung angegeben werden. Datenherausgeber müssen im Open-Data-Portal Schleswig-Holstein nur einmalig erfasst werden und können bei nachfolgenden Veröffentlichungen immer wieder verwendet werden. Von einem neuen Herausgeber werden folgende Informationen benötigt: Name, Logo und kurze Beschreibung

Beispiel:



Abbildung 5.7: Landesamt für Denkmalpflege: Das Landesamt prüft anhand der gesetzlich vorgegebenen Kriterien den besonderen Wert eines Kulturdenkmals und legt die Maßstäbe, die Methodik für die Erfassung und Pflege sowie den Schutzzumfang der Kulturdenkmale fest.

Mit diesen Informationen legt die Open-Data-Leitstelle eine Seite im Open-Data-Portal Schleswig-Holstein für den Herausgeber an. Zunächst in einem Testsystem, nach Rücksprache und gegebenenfalls einer Freigabe durch den Herausgeber dann auch im Produktivsystem. Auf der Herausgeber-Seite können alle von dieser Stelle veröffentlichten Datensätze gefunden werden.

5.3.6 Raumbezug

Zu qualitativ hochwertigen Metadaten gehört auch die Beschreibung, auf welchen geographischen Raum sich ein Datensatz bezieht. Bei Datensätzen der öffentlichen Verwaltung ist dies für gewöhnlich ein verwaltungspolitischer Geo-Bezug (z. B. Gemeinde, Amt oder Kreis). Hierzu gibt es eine Liste von vorgefertigten Einträgen, die über einen URI identifiziert werden.

Die Kennung für das gesamte Bundesland Schleswig-Holstein lautet <http://dcat-ap.de/def/politicalGeocoding/stateKey/01>

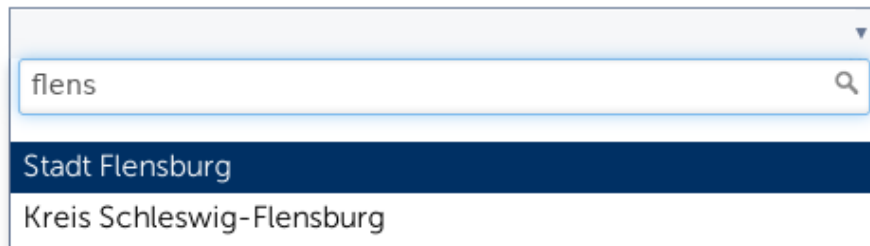
Die Kennungen für die Kreise und kreisfreien Städte in Schleswig-Holstein finden sich hier: <https://www.dcat-ap.de/def/politicalGeocoding/districtKey/>

Die URIs für die Gemeinden sind hier zu finden: <https://www.dcat-ap.de/def/politicalGeocoding/municipalityKey/>

Die URIs für die Ämter sind hier zu finden: <https://www.dcat-ap.de/def/politicalGeocoding/municipalAssociationKey/>

Im Open-Data-Portal Schleswig-Holstein können Sie die jeweilige Kennung unkompliziert über ein Drop-Down-Menü mit automatischer Vervollständigung auswählen und damit den Raumbezug des Datensatzes festlegen:

Raumbezug:



The image shows a search input field with the text 'flens' and a magnifying glass icon. Below the input field, a dropdown menu is open, displaying two options: 'Stadt Flensburg' (highlighted in dark blue) and 'Kreis Schleswig-Flensburg'.

Abbildung 5.8: Screenshot des Drop-Down-Menüs zum Raumbezug im Open-Data-Portal

5.3.7 Veröffentlichungszeitpunkt

Dieses Datum bezeichnet den Zeitpunkt der erstmaligen Veröffentlichung der eigentlichen Daten. Dies bezieht sich nicht auf den Eintrag im Datenregister – wenn die Daten bereits früher auf anderem Wege veröffentlicht wurden, kann das Veröffentlichungsdatum auch früher liegen. Der Veröffentlichungszeitpunkt im Datenregister wird automatisch vermerkt.⁹

Die Information „Zuletzt bearbeitet“ gibt an, wann die Metadaten eines veröffentlichten Datensatzes das letzte Mal verändert wurden. Dies bezieht sich also nicht auf die eigentlichen Daten.

5.3.8 Zeitlicher Bezug

Der zeitliche Bezug eines Datensatzes ist ebenfalls ein wichtiges Auswahlkriterium für Nutzende und darüber hinaus wichtig bei der Suche nach Datensätzen. Zumindest das Startdatum (der Zeitpunkt, für den die ältesten Daten in der Datei enthalten sind) muss angegeben sein. Es sollte nach Möglichkeit auch ein Enddatum (der Zeitpunkt, für den die jüngsten Daten in der Datei enthalten sind) angegeben werden. Es gibt Fälle, bei denen (noch) kein Enddatum existiert (z. B. Pläne, die bis auf Widerruf gelten) oder APIs, die aktuelle Daten liefern. Hier ist die Angabe des Startdatums entsprechend ausreichend.

Erläuterung API

Eine API (Application Programming Interface) ist ein Satz von Befehlen, Funktionen, Protokollen und Objekten, die Programmierer verwenden können, um eine Software zu erstellen oder mit einem externen System zu interagieren. Eine API – auch Programmierstelle genannt – ermöglicht es demnach Anwendungen miteinander zu kommunizieren.

In dem folgenden Schaubild wird der Zusammenhang zwischen dem Veröffentlichungszeitpunkt, der Information „Zuletzt bearbeitet“ sowie dem zeitlichen Bezug noch einmal verdeutlicht:

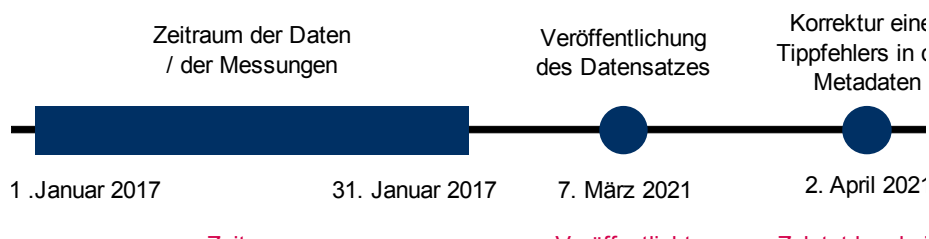


Abbildung 5.9: Übersicht zeitlicher Bezug von Daten

5.3.9 Schlagworte

Damit einzelne Datensätze von den Nutzerinnen und Nutzern in der Fülle des Open-Data-Portals Schleswig-Holsteins auch gefunden werden können, wird jeder Datensatz mit Schlagworten versehen. Die Schlagworte dienen dazu, den Inhalt jedes Datensatzes zu beschreiben und schnell erfassbar zu machen. Gleichzeitig ermöglichen Schlagworte es den Nutzenden, einfach unterschiedliche Datensätze zu ähnlichen Themen zu finden.

Zum Beispiel wird der Datensatz zu der Denkmalliste Kreis Plön mit den Schlagworten „Denkmal, Denkmalliste, Kulturdenkmal“ beschrieben, die wiederum mit anderen Datensätzen zu diesem Thema verknüpft sind. So können über das Schlagwort „Denkmal“ eine Vielzahl

weiterer Denkmallisten aus Schleswig-Holstein gefunden werden.

Empfehlung: Verwendung von Schlagworten

Schlagworte können sich beispielsweise auf die folgenden Informationen beziehen:

- Übergeordnetes Thema des Datensatzes
- Teilaspekte des Datensatzes oder einzelne Themenbereiche
- Orte oder Informationen zu Gebietskörperschaften
- Gesetze und Richtlinien Vermieden werden sollten Schlagworte, die sich wiederholen oder redundant sind. Hierzu zählen beispielsweise:
 - „Open Data“
 - „Schleswig-Holstein“
 - Ortsangaben wie „Kiel“ oder „Flensburg“. Diese Informationen werden bei dem Feld „Raumbezug“ bereits erfasst.
 - Zeitangaben wie „2020“, „März 2020“ oder 1. Quartal 2020. Diese Informationen werden bei dem Feld „Zeitlicher Bezug“ bereits erfasst.
- Schlagworte die lediglich den Titel oder die Beschreibung wiederholen.

5.3.10 Metadaten der Datei (Distribution)

Die Dateien, die in Summe einen Datensatz darstellen, werden ebenfalls mit Metadaten versehen. Im Open-Data-Portal Schleswig-Holstein sind dies regelmäßig der Titel der Datei sowie das Dateiformat. Im Folgenden sehen Sie die Metadaten zu der Datei des in Kapitel 4.1 bereits aufgeführten Beispieldatensatzes „Statistik Nachrichtenbroker“.

Dateien:

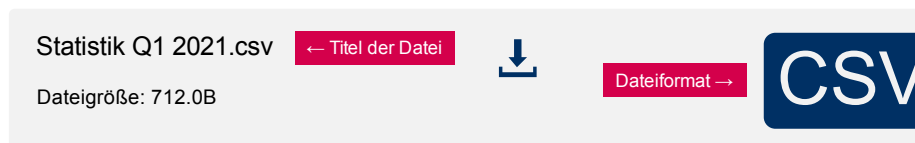


Abbildung 5.10: Bildschirmfoto der Metadaten einer Datei im Open-Data-Portal

5.3.10.1 Titel der Datei

Der Titel beschreibt knapp und möglichst aussagekräftig die eigentliche Datei. Bei dem in Kapitel 4.1 aufgeführten Beispieldatensatz lautet der Titel der CSV-Datei beispielsweise.: „Statistik Q1 2021“.

5.3.10.2 Dateiformate

Das Feld Dateiformate enthält Informationen darüber, in welchen Formaten Daten bereitgestellt wurden und heruntergeladen werden können. Das Feld Dateiformate ist wichtig für die Filterfunktion des Open-Data-Portal Schleswig-Holstein, da hier eine Filterung nach Dateiformaten vorgenommen werden kann.

Kapitel 6

Lizenzen

In diesem Kapitel erfahren Sie, welchen Zweck Lizenzen erfüllen, welche Prüfpflichten vor einer Lizenzauswahl bestehen und schließlich, welche Lizenz (CC BY 4.0) Sie bei einer Veröffentlichung von Daten als Open Data auswählen sollten. Das Kapitel schließt die vorbereitenden Schritte und Rahmenbedingungen der Datenbereitstellung im Open-Data-Portal Schleswig-Holstein ab.

6.1 Was sind Lizenzen?

Die Möglichkeit der freien Weiterverwendung von Daten ist eines der Grundprinzipien von Open Data. Es ist wichtig sicherzustellen, dass veröffentlichte Daten auch durch jede Person für verschiedenste Zwecke uneingeschränkt verwendet werden können. Dazu ist es erforderlich, Daten bei ihrer Veröffentlichung mit Lizenzen zu versehen, die genau dies in rechtlicher Hinsicht ermöglichen. Die Auswahl einer geeigneten Lizenz ist hierbei für jeden zu veröffentlichenden Datensatz individuell vorzunehmen.

Bei der Veröffentlichung von Daten legen Nutzungsbestimmungen fest, unter welchen Bedingungen bzw. mit welchen Einschränkungen ein Datensatz verwendet werden darf. Über die Nutzungsbestimmungen entscheiden Sie als Datenbereitstellerinnen und -bereitsteller (im Rahmen der für Ihre Behörde oder öffentlichen Stelle geltenden Vorgaben) selbst und legen diese mit Hilfe einer Lizenz fest.

Ist keine Lizenz angegeben, dürfen veröffentlichte Datensätze nur sehr eingeschränkt genutzt werden, was den Grundprinzipien von Open Data widersprechen würde. Als grobe Faustregel können Sie sich merken:

Was nicht ausdrücklich erlaubt ist, ist in der Regel verboten!

Die Auswahl und Festlegung einer Lizenz erfolgt auf Ebene eines Datensatzes und gilt dann für alle zugehörigen Ressourcen. Die Lizenz-

angabe muss DCAT-AP.de-kompatibel als URI (Uniform Resource Identifier) erfolgen. Eine Übersicht der mit dem deutschen Metadatenmodell DCAT-AP.de (für weitere Informationen siehe Kapitel 4.2) kompatiblen Lizenzen findet sich hier: <https://www.dcat-ap.de/def/licenses/>.

Welche dieser Lizenzen für eine Veröffentlichung im Open-Data-Portal Schleswig-Holstein empfohlen werden, wird im weiteren Verlauf dieses Kapitels erläutert.

6.2 Prüfpflichten vor der Lizenzauswahl

Bevor Sie sich im Bereitstellungsprozess eines Datensatzes mit der Auswahl einer geeigneten Lizenz auseinandersetzen, muss zunächst geprüft werden, ob überhaupt ein Schutzrecht nach dem Gesetz über Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetz - UrhG) vorliegt.

Dies ist nur dann der Fall, wenn die sogenannte Schöpfungshöhe eines Datensatzes überschritten wird oder es sich um eine Datenbank gemäß § 87a Abs. 1 S. 1 UrhG (eine Sammlung von Werken, Daten oder anderen unabhängigen Elementen) handelt. Die Schöpfungshöhe ist grundsätzlich sehr niedrig: so stellt bereits die Aufnahme eines Fotos mit einem Smartphone ein Werk dar, das mit einer Lizenz versehen werden müsste. Bei einfachen Daten die Tatsachen beschreiben, kann die Schöpfungshöhe jedoch mitunter nicht erreicht sein.

Ergebnis: Besteht kein urheberrechtlicher Schutz, so ist ein Datensatz als gemeinfrei (URI: <http://dcat-ap.de/def/licenses/ccpdm/1.0>) zu kennzeichnen.

Ist die Schöpfungshöhe jedoch erreicht, ist als nächstes zu prüfen, ob es sich um ein amtliches Werk gemäß § 5 Abs. 1 UrhG handelt. Nach § 5 Abs. 1 UrhG genießen „Gesetze, Verordnungen, amtliche Erlasse und Bekanntmachungen sowie Entscheidungen und amtlich verfaßte Leitsätze zu Entscheidungen“ keinen urheberrechtlichen Schutz.

Entscheidend für die Einordnung als amtliches Werk ist dessen Erschaffung im Rahmen einer öffentlich-rechtlichen Handlung, z. B. die Erstellung eines gesetzlich angeordneten Tätigkeitsberichts einer Behörde. Daher können auch Werke Privater, die in Erfüllung einer ihnen übertragenen öffentlichen Aufgabe (Beleihung) erstellt werden müssen, amtliche Werke darstellen.

Ergebnis: Handelt es sich bei einem Datensatz um ein amtliches Werk, so ist es als solches zu kennzeichnen (URI:).

Konnten beide vorherigen Fragen zur Schöpfungshöhe und zum Vorliegen eines amtlichen Werkes verneint werden, ist eine Lizenz auszuwählen. Welche Lizenz konkret auszuwählen ist, wird im nachfolgenden Kapitel erläutert.

6.3 Lizenzauswahl: Creative-Commons-Lizenzen

6.4 Lizenz-Empfehlung für Schleswig-Holstein: CC BY 4.0

Für die Bereitstellung von Daten als offene Daten im Open-Data-Portal Schleswig-Holstein wird durch das Zentrale IT-Management Schleswig-Holstein (ZIT) die folgende Creative-Commons-Lizenz offiziell empfohlen:

Der vollständige Text der Lizenz ist hier zu finden: <https://creativecommons.org/licenses/by-nc/4.0/deed.de>

Der DCAT-AP.de URI dieser Lizenz lautet <http://dcat-ap.de/def/licenses/by-nc/4.0>

Wie genau die Nutzerin oder der Nutzer des Datensatzes dessen Herkunft nennen muss, können Sie als Datenherausgeberin oder -herausgeber festlegen. So kann das Landesamt für Vermessung und Geoinformation Schleswig-Holstein z. B. selbst festlegen, ob es als

- Landesamt für Vermessung und Geoinformation Schleswig-Holstein
- Landesamt für Vermessung und Geoinformation SH,
- LVermGeo Schleswig-Holstein
- LVermGeo-SH

oder mit einer ganz anderen Schreibweise genannt werden möchte.

Diese Lizenz bietet einerseits die gewünschte Rechtssicherheit für die Nutzerinnen und Nutzer der Daten, andererseits ist ihre Verwendung durch Sie als Datenherausgeberin oder -herausgeber mit wenig zusätzlichem Arbeitsaufwand verbunden, weshalb sie stets verwendet werden sollte.

Hinweis:

In der Regel sollte dringend auch eine kommerzielle Nutzung von Daten gestattet werden, da ansonsten deren Verwendung eingeschränkt ist, diese Lizenz ist daher nur in Ausnahmefällen zu verwenden.

Hinweis zur Datenlizenz Deutschland Namensnennung 2.0

Die von vielen öffentlichen Stellen in Deutschland verwendete Datenlizenz Deutschland Namensnennung 2.0 ist problematisch, da sie nicht mit anderen international verwendeten Open-Data-Lizenzen harmonisiert ist und an vielen Stellen offene Fragen lässt. Diese Lizenz

sollte daher für eine Bereitstellung von Daten im Open-Data-Portal Schleswig-Holstein nicht verwendet werden.

Kapitel 7

Bereitstellung von Daten im Open-Data-Portal Schleswig-Holstein

In diesem Kapitel erfahren Sie, wie die Ihnen zur Verfügung stehenden und für eine Veröffentlichung vorgesehenen Daten sowie die dazugehörigen Metadaten in das Open-Data-Portal Schleswig-Holstein gelangen können. Neben der technischen Bereitstellung von Daten und Metadaten werden ebenfalls die Möglichkeiten der Änderung von bereits veröffentlichten Daten erläutert. Nach Abschluss dieses Kapitels sollte das benötigte Wissen für eine Datenbereitstellung im Open-Data-Portal Schleswig-Holstein vorliegen.

7.1 Technische Bereitstellung von Daten

Sowohl für die zu veröffentlichenden Daten selbst, als auch die dazugehörigen Metadaten gibt es unterschiedliche Wege in das Open-Data-Portal. Diese werden im nachfolgenden Kapitel 6.1.1 dargestellt. Das Kapitel 6.1.2 enthält daran anschließend praktische Beispiele für die unterschiedlichen Wege.

7.2 Wege von Daten und Metadaten in das Portal

Mit Daten sind in der Regel eine oder mehrere Dateien gemeint; es ist jedoch auch möglich, auf Dienste mit offenen Schnittstellen (z. B. Web

Feature Service) zu verweisen. Sofern möglich, sollte zu einem Dienst auch immer eine Downloadmöglichkeit der Daten angeboten werden.

Die Metadaten müssen für einen Transport in das Open-Data-Portal als DCAT-AP.de-konformes RDF-Dokument (Resource Description Framework) vorliegen.

Die unterschiedlichen Wege in das Open-Data-Portal (hier als Bereitstellungsmethoden bezeichnet) werden nachfolgend aufgelistet und zur Veranschaulichung grafisch dargestellt.

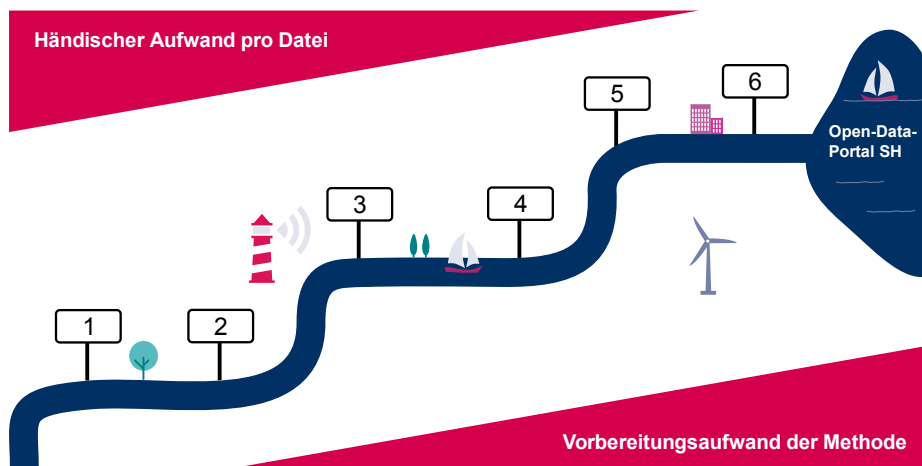


Abbildung 7.1: Der händische Aufwand pro Datei nimmt ab, je größer der Vorbereitungsaufwand der Methode ist

1. Manueller Upload eines Datensatzes in das Open-Data-Portal,
2. Zugriff auf einen vom Internet aus erreichbaren Ablageort durch ein Programm der Open-Data-Leitstelle sowie manueller Upload der Metadaten in das Open-Data-Portal,
3. Zugriff auf eine DCAT-AP.de-konforme RDF-Datei, die von Hand gepflegt wird,
4. ein Fachverfahren, das gelegentlich oder regelmäßig eine DCAT-AP.de-konforme RDF-Datei erzeugt,
5. ein Fachverfahren, das dynamisch eine DCAT-AP.de-konforme RDF-Datei erzeugt,
6. ein Fachverfahren, das per CKAN-API Einträge im Open-Data-Portal anlegen kann.

Der händische Aufwand pro Datei nimmt mit jeder dieser Möglichkeiten ab, dafür nimmt der Aufwand bei der Vorbereitung zu.

Welche Bereitstellungsmethode sich in Ihrem Fall anbieten, hängt von verschiedenen Faktoren ab. Wenden Sie sich bei hier Unterstützungsbedarf bitte an die Open-Data-Leitstelle per E-Mail an das Funktionspostfach opendata@lr.landsh.de.

7.2.1 Welcher Weg sollte gewählt werden?

In Abhängigkeit davon, ob, bzw. wie Ihre Daten auf Ihren Systemen gespeichert sind, fällt die Entscheidung für einen Weg der Daten in das Portal unterschiedlich aus. Die verschiedenen Konstellationen sind:

1. Ich speichere meine Dateien selbst.
 - Bei wenigen Dateien kann eine manuelle Anlage der Einträge im Open-Data-Portal sinnvoll sein.
 - Eine große Anzahl von Dateien könnte über eine DCAT-AP.de-konforme RDF-Datei (von Hand geschrieben oder automatisch generiert) gepflegt werden, die vom Open-Data-Portal regelmäßig gelesen wird.
 - Bei häufigen Änderungen lässt sich mit Hilfe des CKAN-API die Anlage der Datensätze im Open-Data-Portal automatisieren. Damit lässt sich eine zeitnahe Veröffentlichung im Open-Data-Portal realisieren.
2. Ich speichere meine Dateien selber, überschreibe sie aber gelegentlich oder regelmäßig.
 - Für diesen Fall sollten Daten an einem aus dem Internet erreichbaren Ort abgelegt werden, zu dem das Open-Data-Portal eine Verbindung herstellen kann. Die Daten verbleiben also bei ihrem Herausgeber.
 - Eine Veränderung der Daten ist direkt an dem Ablageort möglich. Nach einer abgeschlossenen Veränderung wird die aktualisierte Version der Daten automatisch an das Open-Data-Portal übertragen.
3. Ich möchte selbst keine Dateien speichern, die Dateien sollen im Open-Data-Portal gespeichert werden.
 - Bei wenigen Dateien kann hier ein manuelles Hochladen sinnvoll sein.
 - Kommen oft Dateien hinzu, lässt sich mit Hilfe des CKAN-API das Hochladen und die Anlage der Datensätze im Open-Data-Portal automatisieren.
4. Ich biete ein API mit aktuellen Daten an.
 - Wenn es eine kleine Anzahl APIs ist und sich daran selten Änderungen ergeben, kann eine manuelle Anlage des Eintrags im Open-Data-Portal sinnvoll sein.
 - Bei vielen APIs und/oder Änderungen kann eine DCAT-konforme RDF-Datei sinnvoll sein, die vom Open-Data-Portal regelmäßig gelesen wird.

7.3 Praktische Beispiele für die Wege in das Open-Data-Portal

7.3.1 Denkmalliste

Das Landesamt für Denkmalpflege muss regelmäßig die Denkmalliste Schleswig-Holstein veröffentlichen. Die Daten dazu liegen im vom Dataport entwickelten und betriebenen Fachanwendung "DISH". Die bisherige Lösung sah so aus, dass etwa einmal pro Quartal von Hand ein Export von 15 großen PDF-Dateien (eine pro Kreis) angestoßen und heruntergeladen wurde. Diese 15 PDF-Dateien mussten anschließend im Landesportal Schleswig-Holstein hochgeladen und verlinkt werden.

Die Anbindung an das Open-Data-Portal ist so gestaltet, dass der PDF-Export und ein zusätzlicher Export als JSON-Datei wöchentlich automatisch vom Fachverfahren durchgeführt wird. Die Dabei erzeugten Dateien landen in einem per HTTP erreichbaren Verzeichnis. Dort werden sie vom Open-Data-Portal eingesammelt und weiter verarbeitet. Zusätzlich zur JSON-Datei wird dabei die [Denkmalliste im CSV-Format](#) erzeugt, da dies für viele Nutzer*innen leichter zu interpretieren ist. Die [große JSON-Datei](#) wird zusätzlich kreisweise zerlegt, um - passend zu den PDF-Dateien - Denkmalliste pro Kreis anbieten zu können, z.B. zum [Kreis Rendsburg-Eckernförde](#). Automatisch werden Datensätze im Open-Data-Portal angelegt und in Zeitreihen eingeordnet, um so persistente und zitierfähige Adressen für die Denkmalliste(n) zu schaffen.

Als zusätzlichen Bonus werden nun auch Fotos der Denkmäler veröffentlicht. Auch dabei wurde darauf geachtet, möglichst wenig Arbeitsaufwand zu produzieren. Die Fotos werden in einem Netzlaufwerk abgelegt. Da im Dateinamen die eindeutige Kennung des Denkmals enthalten ist, kann ein automatischer Abgleich mit der Denkmalliste erfolgen und die JSON- und CSV-Dateien entsprechend angereichert werden.

7.3.2 Badegewässerqualität

Die Daten zur Badegewässerqualität werden vom Ministerium für Soziales, Gesundheit, Jugend, Familie und Senioren bereitgestellt. Sie bestehen aus fünf Datensätzen:

1. [Stammdaten](#),
2. [Informationen zur vorhandenen Infrastruktur](#),
3. [Einstufung der Badegewässerqualität](#),
4. [Saisondauer](#) und
5. [Messungen](#).

Aus der Fachanwendung zur Kontrolle der Badegewässerqualität werden diese Datensätze täglich als fünf CSV-Dateien exportiert und in einem per Internet erreichbaren Verzeichnis abgelegt. Das Open-Data-Portal prüft täglich, ob es Änderungen an den Dateien gegeben hat.

Ist dies der Fall, werden die Datensätze im Open-Data-Portal entsprechend angepasst. Da an die CSV-Dateien für die Datensätze 3 bis 5 lediglich neue Werte am Ende hinzugefügt werden, muss hier nur das Änderungsdatum aktualisiert werden. Bei den Datensätze 1 und 2 werden die Datensätze archiviert und in Zeitreihen eingeordnet, so dass man später Entwicklungen (z. B. beim Ausbau der Infrastruktur an Badestellen) nachvollziehen kann.

7.4 Änderungen an veröffentlichten Datensätzen

Mitunter werden Änderungen an veröffentlichten Datensätzen erforderlich. Dies ist der Fall, wenn Datensätze Fehler enthalten, oder sie einer regelmäßigen Aktualisierung bedürfen. Mit beiden Optionen ist unterschiedlich umzugehen.

7.4.1 Korrektur von Fehlern in Datensätzen

Sofern Fehler in Datensätzen bestehen, können diese durch eine erneute Bearbeitung im Open-Data-Portal durch Sie als Datenherausgeberin oder -herausgeber behoben werden.

Als Beispiel für eine Korrektur von Fehlern sei der Datensatz Ausführung des Landes Schleswig-Holstein 4. Quartal 2017 genannt. Die hier ursprünglich verlinkte Distribution G_III_1_vj174_SH.xlsx wurde gelöscht und durch eine korrigierte Version G_III_1_vj174_SH_korr.xlsx ersetzt. Der Hinweis auf die Korrektur einer Distribution ist (wie hier erfolgt) stets hervorzuheben.

7.4.2 Reguläre Aktualisierung der Daten

Es gibt Datensätze, die einer regelmäßigen Aktualisierung bedürfen, und deshalb wiederkehrend verändert werden. Hierzu gibt es zwei unterschiedliche Formen der Aktualisierung:

Reguläre Aktualisierung von Daten

- Die bisherigen Daten werden durch einen neuen Stand der Daten überschrieben.
- An die bisherigen Daten werden die neuen Daten angehängt.

Beim Überschreiben der Daten wird ein neuer Datensatz erzeugt. Der bisherige Datensatz bekommt ein Enddatum zugewiesen, der neue Datensatz bekommt dieses Datum als Startdatum. Das Datum der Aktualisierung kann nicht unbedingt aus der Datei selbst entnommen werden, da diese möglicherweise keine Zeitangaben enthält. Ein Beispiel für diese regelmäßige Überschreibung ist der Datensatz zur Badegewässer Infrastruktur, bei dem immer wieder eine neue Version verfügbar gemacht wird.

Beim Anhängen müssen nur der Gültigkeitszeitraum und das Änderungsdatum in den Metadaten des bestehenden Datensatzes aktualisiert werden. Ist bekannt, an welcher Stelle in der Datei das Datum der „Messung“ angegeben ist, kann man das neue Enddatum des Gültigkeitszeitraums automatisch ermitteln. Ein Beispiel für diese Art der Aktualisierung ist der Datensatz zu den Badegewässer Messungen.

Das Änderungsdatum ist nicht unbedingt übereinstimmend mit dem Ende des Gültigkeitszeitraums. In beiden Fällen kann das Änderungsdatum ohne den/die Datenherausgeber*in bestimmt werden. Es bieten sich zwei Werte an:

- Datum, an dem die Änderung bemerkt wird
- HTTP-Header Last-Modified, der beim Abruf der Datei gesendet wird

Im Folgenden ebenfalls ein Beispiel aus dem Bereich der Badegewässer, bei dem ersichtlich wird, wie sich die unterschiedlichen Arten der Aktualisierung in der Praxis darstellen:

| Datei | Art der Aktualisierung |
|---------------------------|------------------------|
| v_badegewaesser_odata.csv | Überschreiben |
| v_badesaison_odata.csv | Anhängen |
| v_einstufung_odata.csv | Anhängen |
| v_infrastruktur_odata.csv | Überschreiben |
| v_proben_odata.csv | Anhängen |

Kapitel 8

Ende des Open Data Leitfadens für Schleswig-Holstein

Sie befinden sich nun am Ende des Open Data Leitfadens für Schleswig-Holstein. Mit dem in diesem Leitfaden vermittelten Wissen sollten Sie in der Lage sein, Ihre Daten im Open-Data-portal Schleswig-Holstein bereitzustellen. Bei weitergehenden Fragen zum Thema Open Data können Sie sich an ... wenden.

Glossar

API *API* ist die Abkürzung für “Application Programming Interface”. Ein API ist eine Schnittstelle, die eine dynamische Bereitstellung von Daten ermöglicht.

CC → Creative Commons

Creative Commons *Creative Commons* (CC) ist eine Non-Profit-Organisation, die in Form vorgefertigter Lizenzverträge eine Hilfestellung für Urheber zur Freigabe rechtlich geschützter Inhalte anbietet.

Datei Eine Datei ist ein Bestand meist inhaltlich zusammengehöriger Daten, der auf einem Datenträger oder Speichermedium gespeichert ist.

Dateiformat Ein Dateiformat legt den Inhalt von Dateien fest. Es handelt sich dabei um vereinbarte Konventionen der inneren Struktur, wie Informationen eines bestimmten Datentyps, beispielsweise Text, Grafik, Video, Audio, Tabellen angeordnet sind. Für Open Data geeignete Dateiformate sollten nach Möglichkeit offen standardisiert sein. Beispiele für Dateiformate sind CSV, JSON, XML, RDF oder das Excel-Format XLSX.

Daten Daten bilden die einzelnen Werte innerhalb eines Datensatzes. Unter Daten werden im IT-Kontext jede Art von Informationen verstanden, die durch einen Computer interpretierbar sind und weiterverarbeitet werden können.

Datenkatalog Eine Sammlung aller Datensätze, die für eine Veröffentlichung als Open Data identifiziert wurden. In der einfachsten Form kann dies bereits eine Excel-Tabelle sein.

Datenportal Ein Datenportal ist eine Webanwendung, über die in einem Bestand von Datensätzen gesucht und navigiert sowie auf Datensätze zugegriffen werden kann. Jeder Datensatz ist dabei durch Metadaten beschrieben, durch die er besser auffindbar gemacht wird.

Datensatz Eine Ansammlung von inhaltlich zusammenhängenden Daten in einer Datei. Ein Datensatz kann aus einer Vielzahl von Daten bestehen, jedoch auch aus nur einem Datum.

Datenscreening Alternative Bezeichnung: Datenrecherche. Ein strukturiertes Vorgehen zur Identifizierung von Datensätzen zur Veröffentlichung als Open Data.

DCAT-AP.de *DCAT-AP.de* ist die deutsche Adaption des europäischen Metadatenstandards DCAT-AP und wird als gemeinsames deutsches Metadatenmodell zum Austausch von offenen Verwaltungsdaten verwendet.

Distribution Eine Distribution ist die technische Beschreibung einer konkreten Datei.

CKAN *CKAN* ist ein Open Source Datenkatalog, der häufig das Grundgerüst für Open Data Portale bildet. Das Open-Data-Portal Schleswig-Holstein nutzt derzeit CKAN als Katalogsystem.

CSV Das Datei-Format *CSV* steht im englischen Original für „Comma Separated Values“. Hierunter werden tabellarische Daten verstanden, die in einem Textformat gespeichert und über einen Zeichentrenner (meistens Komma oder Strichpunkt) getrennt hintereinandergeschrieben werden.

GovData *GovData* ist das zentrale Metadatenportal in Deutschland. In diesem Portal sind die Metadaten von allen angebotenen Portalen der Bundes-, Landes-, und kommunalen Ebene erfasst.

Harvesting Harvesting bezeichnet das systematische Sammeln (Ernten) von Daten aus Datenbanken, Repositorien oder anderen digitalen Quellen durch eine Software.

JSON JSON ist die Abkürzung für „JavaScript Object Notation“, ein Dateiformat, das eine hohe Maschinenlesbarkeit aufweist.

LOD LOD steht für „Linked Open Data“; miteinander vernetzte offene Daten.

Maschinenlesbarkeit Die Maschinenlesbarkeit eines Datensatzes sagt aus, ob dieser von einer Maschine, sprich Software, gelesen bzw. interpretiert und weiterverarbeitet werden kann.

Metadaten Metadaten sind „Daten über Daten“; sie beschreiben also andere Daten und sorgen dafür, dass diese in Datenportalen gefunden werden können.

Musterdatenkatalog Ein Datenkatalog, der Daten von allen Kommunen in Deutschland beinhaltet, deren offene Daten über das Datenportal für Deutschland, *GovData.de*, auffindbar sind.

Open Data Open Data sind ungefilterte und maschinenlesbare elektronische Daten, die jedem öffentlich, zweckfrei und unverbindlich zur Verfügung gestellt werden.

PSI „Public Sector Information“-Richtlinie der EU, welche die Weiterverwendung von Informationen des öffentlichen Sektors regelt. Sie stellt den europäischen Rechtsrahmen für Open Data dar.

RDF Resource Description Framework (sinngemäß: System zur Beschreibung von Ressourcen). Ein Datenformat zur Beschreibung von Ressourcen im Internet. Metadaten werden in RDF-Form bereitgestellt.

Rohdaten Grundlage von Datensätzen sind unbearbeitete Informationen. Diese werden entweder unmittelbar – in Echtzeit oder Nachhinein – „roh“ veröffentlicht (etwa von einer Messtelle per API/Schnittstelle).

URI *Uniform Resource Identifier* (englisch für „einheitlicher Bezeichner für Ressourcen“). Der Identifikator besteht aus einer Zeichenfolge, die zur Identifizierung einer Ressource (z. B. eine Datei oder eine Website) dient.

URL *Uniform Resource Locator* (englisch für „einheitlicher Ressourcenzeiger“). Eine URL ist die Adresse einer Website.

Quellenverzeichnis

[5starinfo] 5-Sterne Offene Daten, URL: <https://5stardata.info/de/>,
zuletzt abgerufen 2022-04-19

Bildverzeichnis

- **Alle nicht einzeln genannten Grafiken und Screenshots:**
Land Schleswig-Holstein, veröffentlicht unter Creative Commons
Namensnennung 4.0 International (CC BY 4.0)

Impressum

Herausgeber: Land Schleswig-Holstein

Lizenz: Der Text des Handbuchs ist unter einer Creative Commons Namensnennung 4.0 International Lizenz (CC BY 4.0) veröffentlicht. Bilder und andere Elemente, deren Urheberrecht bei Dritten liegen, sind ausgenommen. Quellenverzeichnis und Bildverzeichnis mit entsprechenden Urheberrechtsangaben sind im Handbuch enthalten.

Quelle: Der Quelltext für den Leitfaden befindet sich in folgendem Repository: <https://code.schleswig-holstein.de/opendata/leitfaden>

Stand: 2022-04-20