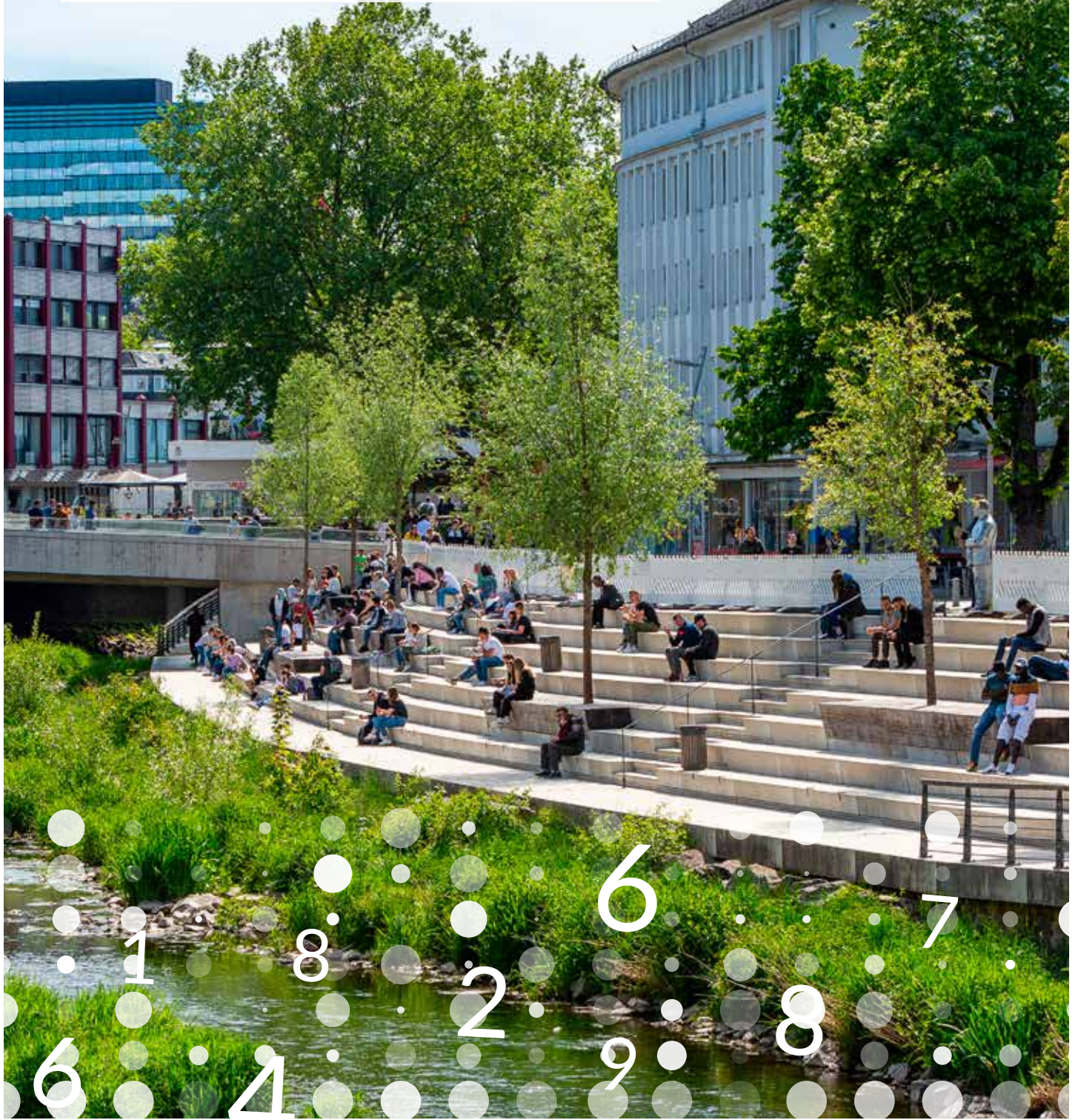


Offene Daten der Zivilgesellschaft

Ein Leitfaden für die Praxis



Impressum

Bertelsmann Stiftung, Gütersloh

November 2023

Herausgeber

Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256
33311 Gütersloh
Telefon +49 5241 81-0
www.bertelsmann-stiftung.de

Verantwortlich

Mario Wiedemann

Autor

Bernhard Krabina

Lektorat

Rudolf Jan Gajdacz, München

Grafikdesign

Nicole Meyerholz, Bielefeld

Bildrechte

Titelfoto: © Harlekin-Graphics – stock.adobe.com

Der **Text** und die **Grafiken** dieser Publikation sind lizenziert unter der Creative Commons Namensnennung 4.0 International (CC BY 4.0) Lizenz. Den vollständigen Lizenztext finden Sie unter: <https://creativecommons.org/licenses/by/4.0/legalcode.de>



Davon ausgenommen sind alle **Fotos** und **Logos**, sie unterfallen nicht der oben genannten CC-Lizenz.

Online ist die Publikation unter <https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/offene-daten-der-zivilgesellschaft> abrufbar.

Zitiervorschlag

Krabina, Bernhard. Offene Daten der Zivilgesellschaft. Ein Leitfaden für die Praxis. Bertelsmann Stiftung. Gütersloh 2023.

DOI 10.11586/2023089

Offene Daten der Zivilgesellschaft

Ein Leitfaden für die Praxis

Bernhard Krabina

Inhalt

Vorwort	5
Das Wichtigste auf einer Seite	6
Warum offene Daten? – Open Commons	7
Wie vorgehen? – Schritte zu offenen Daten	9
Welche Daten? – Datenkatalog	13
Wo veröffentlichen? – Dateninfrastrukturen	17
Wie beginnen? – Phasen der Umsetzung	19
Jetzt starten – eine Checkliste	21
Worauf gefasst sein? – Häufige Fragen und Antworten	22
Welche Begriffe sollten Sie kennen? – Glossar	24
Wo können Sie weiterlesen? – Weiterführende Informationen	25

Vorwort

Die Zivilgesellschaft in Deutschland leistet auf vielfältige Weise einen Beitrag für das Gemeinwohl. Zahlreiche gemeinnützige Vereine, Organisationen und Institutionen engagieren sich in sozialen Bereichen wie Armutsbekämpfung, Bildungs- oder Demokratieförderung. Sie unterstützen die politische Debatte und die Partizipation aller Bürgerinnen und Bürger am gesellschaftlichen Leben. Die Zivilgesellschaft stellt dafür eine breite Palette an Angeboten und Aktivitäten bereit, organisiert Veranstaltungen oder Initiativen, um gesellschaftliche Veränderungen zu initiieren.

Ein Beitrag zur Förderung des Gemeinwohls kann aber auch über eine bisher weniger verbreitete Art und Weise erfolgen: mit offenen Daten, die der Allgemeinheit kostenlos zur Verfügung gestellt werden. Die öffentliche Verwaltung in Deutschland stellt zunehmend offene Daten, auch Open Data genannt, auf Bundes-, Landes- und kommunaler Ebene bereit. Diese Daten werden von unterschiedlichen Zielgruppen, wie z. B. Wirtschaft, Wissenschaft, Medien oder der öffentlichen Verwaltung selbst, genutzt.

In zivilgesellschaftlichen Organisationen ist dies bisher weniger der Fall, bietet aber ein großes Potenzial. In einer Umfrage unter zivilgesellschaftlichen Organisationen haben auf die Frage, welchen Mehrwert sie offenen Daten zuschreiben, 85 Prozent der Befragten mit „Stärkung der Transparenz“ geantwortet. Offene Daten schaffen bessere Bedingungen für evidenzbasierte Interessenvertretung und organisationsintern können sie Wirkung erzeugen, indem Daten einfacher in ganz unterschiedlichen Zusammenhängen genutzt werden können.

Die Bertelsmann Stiftung veröffentlicht bereits einen Teil ihrer Daten, die für Studien und Portale verwendet werden, als Open Data. Mit dem Wegweiser Kommune (wegweiser-kommune.de) betreiben wir beispielsweise ein großes Datenportal, das kommunale Daten unter einer offenen Lizenz und per Schnittstelle (API) zur Verfügung stellt. Wir werden dieses Engagement ausbauen und perspektivisch noch mehr Daten offen zur Verfügung stellen – dazu bauen wir unser eigenes Datenmanagement aus und entwickeln aktuell einen Datenkatalog als Basis für diese strategische Weiterentwicklung.

Der Austausch mit anderen zivilgesellschaftlichen Organisationen hat uns gezeigt, dass das Interesse an Open Data in der Zivilgesellschaft groß ist. Ich wünsche allen Leserinnen und Lesern mit dem vorliegenden Leitfaden gutes Gelingen dabei, mehr Daten offen zur Verfügung zu stellen. Wir freuen uns über einen Austausch mit weiteren zivilgesellschaftlichen Organisationen. Lassen Sie uns gemeinsam einen Beitrag zu einem wachsenden Datenökosystem leisten.

Dr. Ralph Heck

Vorstandsvorsitzender
Bertelsmann Stiftung

Das Wichtigste auf einer Seite

Offene Daten („Open Data“) sind Daten, die der Allgemeinheit kostenlos zur Weiterverwendung zur Verfügung gestellt werden.

Offene Daten stellen eine Basisinfrastruktur für Digitalisierung dar. Ziel ist es, dass die Daten zur Information und Transparenz beitragen und diese Daten möglichst einfach wiederverwendet werden können: von Wirtschaft, Wissenschaft, Medien und öffentlichen Stellen, und daraus beispielsweise neue Anwendungen (Apps) oder Aufbereitungen (Visualisierungen) entstehen.

Was sind denn nun Daten? Technisch gesehen sind auch PDF-Dokumente oder die Internetseiten Ihrer Organisation Daten. Bei offenen Daten sind aber insbesondere strukturierte Daten gemeint, also z. B. Daten aus einer Datenbank oder die tabellarischen Daten, aus der eine Tabelle oder Abbildung im PDF-Dokument erzeugt wurde, oder Daten, die einer Visualisierung auf einer Website zugrunde liegen.

Die Veröffentlichung von offenen Daten erfolgt im Wesentlichen in drei Schritten:

1. **unter Angabe einer freien Lizenz**, die spezifiziert, wie Nutzer:innen die Daten verwenden dürfen (rechtlicher Aspekt),
2. **in offenen und maschinenlesbaren technischen Formaten** (technischer Aspekt),
3. **beschrieben durch Metadaten** (Angaben zu den veröffentlichten Daten, organisatorischer Aspekt).

Offene Daten bieten zivilgesellschaftlichen Organisationen eine große Chance, die eigene Kompetenz im Umgang mit Daten zu erhöhen und mit der freien Veröffentlichung die Orientierung am Gemeinwohl zu stärken.

Der vorliegende Leitfaden für offene Daten enthält neben der Beschreibung dieser wichtigsten Schritte noch weitere Informationen zu Datenkatalogen und Datenportalen. Ebenso sind unterschiedliche Phasen der Umsetzung sowie Fragen und Antworten beschrieben.

Wenn Sie noch nie etwas von DCAT-AP, JSON oder APIs gehört haben, sind Sie hier richtig. Dieser Leitfaden hat zum Ziel, Ihnen kurz und kompakt das Wichtigste über offene Daten zu vermitteln und auf zahlreiche weiterführende Quellen zu verweisen.

Warum offene Daten? – Open Commons

In den vergangenen Jahren haben insbesondere zivilgesellschaftliche Organisationen das Thema der offenen Verwaltungsdaten („Open Government Data“, kurz: OGD) vorangetrieben. Also Daten, die von öffentlichen Stellen gesammelt, erstellt oder bezahlt wurden. Dies hat dazu geführt, dass zahlreiche Behörden Daten veröffentlichen und in Metadatenportalen wie GovData¹ und dem europäischen Datenportal² verzeichnen.

Doch auch in zivilgesellschaftlichen Organisationen wie Nichtregierungsorganisationen (NGOs), Vereinen, Verbänden, Stiftungen oder Gewerkschaften existieren viele Daten, die von großem Nutzen für die Gesellschaft sein können.³ Obwohl diese Organisationen aufgrund ihres Auftrags bzw. Selbstverständnisses gemeinwohlorientiert sind und den Zweck haben, einen Mehrwert für unsere Gesellschaft zu erzeugen, ist die Veröffentlichung offener Daten noch keine weit verbreitete Praxis. In der Regel veröffentlichen zivilgesellschaftliche Organisationen viele Informationen in Publikationen (Print-Publikationen oder PDF-Dokumenten) oder auf Webseiten (HTML-Seiten), ohne jedoch die zugrunde liegenden Daten in strukturierter Form zum Download anzubieten.

Der französische Verein „Open Food Facts“⁴ hatte beispielsweise die Idee, über eine App Daten zu Lebensmitteln (Inhaltsstoffe, Erzeugung, Verpackung) zu sammeln. Dadurch, dass die Daten frei weiterverwendet werden können, sind bereits zahlreiche Anwendungen entstanden, etwa zur Ernährungs- oder Allergiebera-

tung.⁵ Durch die freie Verfügbarkeit der eigens entwickelten Software konnten ähnliche Projekte entstehen, etwa „Open Beauty Facts“⁶ für Kosmetikprodukte.

Frei verfügbare Daten können als „Commons“ verstanden werden, also Ressourcen, die wie ein Gemeingut oder öffentliches Gut bzw. Allmende von allen genutzt werden können.⁷ Mit „Open Commons“ (oder auch „Data Commons“⁸) kann man digitale Gemeingüter bezeichnen, also Wissen, Werke und Inhalte, die einer Gemeinschaft frei und offen zugänglich sind und so von allen ohne große Hindernisse genutzt werden können.⁹ Beispiele sind digitale Videos oder Tonaufnahmen, Datensätze, Baupläne, Landkarten oder Software.

Internationale Charta für offene Daten

Die Sunlight Foundation¹⁰ hat bereits 2010 zehn Prinzipien für offene Daten entwickelt, die als Forderung an staatliche Stellen formuliert sind und immer wieder abgewandelt wurden. Die NGO „Open Data Charter“ (ODC) hat 2015 sechs Prinzipien¹¹ erstellt, die von zahlreichen staatlichen und nicht staatlichen Organisationen freiwillig unterzeichnet wurden. Für zivilgesellschaftliche Organisationen bieten wir in Ab-

1 <https://www.govdata.de> (Download 12.10.2023).

2 <https://data.europa.eu> (Download 12.10.2023).

3 Siehe dazu auch Juech 2022.

4 <https://world.openfoodfacts.org/> (Download 12.10.2023).

5 <https://www.data.gouv.fr/en/datasets/open-food-facts-produits-alimentaires-ingredients-nutrition-labels/> (Download 12.10.2023).

6 <https://world.openbeautyfacts.org/> (Download 12.10.2023).

7 <https://openall.info/open-culture-offene-kultur/open-knowledge/commons/> (Download 12.10.2023).

8 <https://www.givingtuesday.org/data-commons/> (Download 12.10.2023).

9 <https://opencommons.linz.at/vision/> (Download 12.10.2023).

10 <https://sunlightfoundation.com/policy/documents/ten-open-data-principles/> (Download 12.10.2023).

11 <https://opendatacharter.net/principles/> (Download 12.10.2023).

ABBILDUNG 1 Open Data Charter

1. Standardmäßig offen („Open by Default“)	Die Organisation unterzieht sich einem Wandel hin zu der Einstellung, dass alles veröffentlicht werden sollte und Daten bzw. Informationen nur zurückgehalten werden, sofern wichtige Gründe (z. B. Datenschutz oder Sicherheit) dagegensprechen.
2. Zeitnah und umfassend	Offene Daten sind nur dann wertvoll, wenn sie noch relevant sind. Die schnelle und umfassende Veröffentlichung von Informationen ist entscheidend für ihren Erfolg. Die Organisation sollte Daten so vollständig wie möglich in ihrer ursprünglichen, unveränderten Form bereitstellen.
3. Zugreifbar und nutzbar	Wenn sichergestellt ist, dass die Daten maschinenlesbar und leicht auffindbar sind, werden sie weiterverbreitet. Portale sind eine Möglichkeit, dies zu erreichen. Es ist aber auch wichtig, sich Gedanken über die Benutzerfreundlichkeit beim Zugriff auf Daten zu machen, einschließlich der Dateiformate, in denen die Informationen bereitgestellt werden. Die Daten sollten kostenlos und unter einer offenen Lizenz zur Verfügung gestellt werden, z. B. unter den von Creative Commons entwickelten Lizenzen.
4. Vergleichbar und interoperabel	Daten haben einen Multiplikatoreffekt. Je mehr hochwertige Datensätze zur Verfügung stehen und je einfacher sie miteinander kombiniert werden können, desto größer ist der potenzielle Wert, den man aus ihnen ziehen kann.
5. Für eine bessere Regierungsführung und mehr Bürgerbeteiligung	Offene Daten können den Bürger:innen und Regierungen eine bessere Vorstellung davon vermitteln, was die Organisation macht. Diese Transparenz kann die Dienstleistungen verbessern und dazu beitragen, die Organisation zur Rechenschaft zu ziehen.
6. Für inklusive Entwicklung und Innovation	Schließlich können offene Daten eine inklusive wirtschaftliche Entwicklung vorantreiben. Ein besserer Zugang zu Daten kann z. B. die Landwirtschaft effizienter machen oder zur Bekämpfung des Klimawandels eingesetzt werden. Schließlich denken wir bei offenen Daten oft nur an die Verbesserung der Leistung von Regierungen, aber es gibt ein ganzes Universum von Unternehmern, die mit offenen Daten Geld verdienen.

Quelle: Eigene Darstellung. In Anlehnung an opendatacharter.net.

| BertelsmannStiftung

Abbildung 1 eine abgewandelte, freie Übersetzung der Prinzipien.

Organisationen, die diese Prinzipien unterstützen, können eine Erklärung unterzeichnen und auch auf der ODC-Website angeführt werden.¹²

Für zivilgesellschaftliche Organisationen ist es daher aus vielen Gründen vorteilhaft, sich mit offenen Daten auseinanderzusetzen:

- Der Ansatz „Open by Default“ maximiert das Potenzial zur Weiternutzung von Daten und Informationen. Freie Lizenzen und offene, strukturierte Formate minimieren die Barrieren zur Nutzung.
- Ein Angebot an Daten und Informationen stärkt die Gemeinwohlorientierung und unterstützt die Ziele der Organisation.
- Die Weiternutzung steigert die Bekanntheit der Organisation/der Projekte und kann zu einem zusätzlichen gesellschaftlichen Mehrwert führen.
- Der Fokus auf Daten fördert die Datenkompetenz in der eigenen Organisation und deren Mitarbeiter:innen.
- Die Veröffentlichung trägt zur Verbesserung der Datenqualität bei (durch vorbereitende Maßnahmen, Metadaten und Feedback durch Nutzer:innen).
- Mit einem Angebot an offenen Daten werden viele Stakeholder bedient: Datenjournalist:innen, Forscher:innen, Bürger:innen sowie andere Organisationen.
- Durch datenbasierte Angebote wird auch die Datenkompetenz in der Gesellschaft erhöht.
- Komplexe gesellschaftliche Herausforderungen benötigen die enge Zusammenarbeit zwischen Staat, Zivilgesellschaft und Privatwirtschaft einerseits sowie eine breite Verfügbarkeit von frei verfügbaren Daten andererseits.

¹² <https://opendatacharter.net/endorsing-institutions/> (Download 12.10.2023).

Wie vorgehen? – Schritte zu offenen Daten

Offene Daten können ein komplexes Thema sein, wenn der Anspruch besteht, jedes kleinste Detail nachzuvollziehen. Dabei sind nur drei wesentliche Schritte nötig, um Daten zu öffnen.

Schritt 1: Freie Lizenzen (rechtlicher Aspekt)





Bei einer Veröffentlichung von Daten und Informationen auf Webseiten wird i. d. R. keine Lizenz der Inhalte vergeben. Das bedeutet, dass die Inhalte zwar gelesen, heruntergeladen und für den eigenen Gebrauch genutzt, aber nicht weiterverwendet werden dürfen. Dies gilt auch implizit selbst dann, wenn sich keine diesbe-

züglichen Angaben im Impressum finden. Häufig sind in diesem sogar noch weitere Einschränkungen vorgesehen, im Sinne von „Alle Rechte vorbehalten“.

Da eines der Ziele von offenen Daten deren Wiederverwendung ist, ist für deren Nutzung eine explizite Einräumung einer freien Lizenz gemäß der Open Definition¹³ bzw. eine deutliche Kennzeichnung von Nutzungsbedingungen erforderlich. Es ist daher (generell oder pro Datensatz) zu entscheiden, welche Bedingungen anwendbar sind, wobei der Grundsatz gelten sollte: **So offen wie möglich!**

¹³ <https://opendefinition.org/> (Download 12.10.2023).

ABBILDUNG 2 Varianten von Lizenzen

Lizenzvariante	Anmerkung
gemeinfrei 	Geistige Schöpfungen, an denen keine Immaterialgüterrechte, insbesondere kein Urheberrecht bestehen, gelten als gemeinfrei. ¹⁴ Darunter fallen auch Werke ohne Schöpfungshöhe. Die Public Domain Mark ist keine Lizenz, sondern markiert Inhalte, die keine Lizenz benötigen. https://creativecommons.org/publicdomain/mark/1.0/deed.de
Zero (CC0) Empfehlung 	Zero ist die freieste Lizenz. Sie bedeutet, dass die Verwendung der Daten an keinerlei Bedingungen geknüpft ist. Sie können diese Lizenz auch verwenden, wenn Sie nicht wissen, ob die Inhalte gemeinfrei sein müssten. https://creativecommons.org/publicdomain/zero/1.0/deed.de
Namensnennung (CC BY) 	Namensnennung bedeutet, dass die Verwendung an die einzige Bedingung geknüpft ist, die Quelle zu nennen. https://creativecommons.org/licenses/by/4.0/deed.de
andere Lizenzen 	Es existieren noch zahlreiche weitere Lizenzen. Machen Sie davon nur Gebrauch, falls unbedingt nötig. Weitere Einschränkungen oder „exotischere“ Lizenzvarianten erschweren die Wiederverwendung bzw. Verknüpfung mit anders lizenzierten Daten.

Quelle: Eigene Darstellung.

| BertelsmannStiftung

¹⁴ Da die Einschätzung, was gemeinfrei ist (was z. B. als „freies Werk“ nach § 5 Abs. 1 UrhG gilt bzw. ob die Daten überhaupt „Schöpfungshöhe“ oder Werkcharakter erreichen), nicht immer einfach ist, kann es in der Praxis vorkommen, dass eigentlich gemeinfreie Inhalte trotzdem als „Zero“ gekennzeichnet werden. Das mag zwar juristisch nicht korrekt sein, ist aber durchaus im Sinne der Nutzer:innen, da sie sich bei „Zero“ darauf verlassen können, die Daten uneingeschränkt nutzen zu können.

Folgende Lizenzvarianten sind darüber hinaus noch erwähnenswert:

- **Open Data Commons¹⁵** ist eine Reihe von rechtlichen Werkzeugen und Lizenzen, die insbesondere für die Lizenzierung von Datenbanken geschaffen wurden. Sie sind weniger weit verbreitet als CC-Lizenzen. Die PDDL (Open Data Commons Public Domain Dedication Licence) entspricht etwa CC0, die ODC-By (Open Data Commons Attribution License) entspricht etwa CC-BY. Vorsicht geboten ist bei der ODbL (Open Data Commons Open Database License). Sie enthält neben der Namensnennung auch eine „Share-Alike“-Klausel, entspricht daher eher CC-BY-SA. „Share-Alike“ bedeutet Weitergabe nur unter gleichen Bedingungen, was die Weiternutzung der Daten erschwert und deshalb nur in Ausnahmefällen angewandt werden sollte.
- **Datenlizenz Deutschland¹⁶** ist eine speziell für den öffentlichen Sektor in Deutschland angepasste offene Lizenz in den Ausprägungen „Zero“ und „Namensnennung“. Für zivilgesellschaftliche Organisationen sind die international bekannteren CC-Lizenzen oder die Open-Data-Commons-Lizenzen zu bevorzugen.

Bei zivilgesellschaftlichen Organisationen fällt die von Behörden oft verwendete Begründung weg, dass die Erhebung der Daten bereits mit Steuergeld bezahlt wurde und daher eine möglichst freie (CC0- oder CC-BY-) Lizenz gewählt werden sollte. Man könnte daher verleitet sein, an einschränkenderen Lizenzvarianten wie „Share-Alike“ (Weitergabe nur zu gleichen Bedingungen) oder „NonCommercial“ (Verwendung nur zu nicht kommerziellen Zwecken) Gefallen zu finden. Diese Lizenzvarianten sind aber nicht empfehlenswert, da insbesondere die nicht kommerziellen Zwecke sehr einschränkend ausgelegt werden und damit nur noch eine Nutzung zum rein privaten Gebrauch übrig bleibt.¹⁷ Diese stärker einschränkenden Lizenzen widersprechen daher dem Ziel von offenen Daten, eine möglichst breite Weiternutzung zu ermöglichen.

15 <https://opendatacommons.org/> (Download 12.10.2023).

16 <https://www.govdata.de/web/guest/lizenzen> (Download 12.10.2023).

17 siehe z. B. Urteil des Landesgerichts Köln: <https://www.urheberrecht.org/news/5162/>.

TIPP: Freie Lizenzen auch für Dokumente!

Obwohl hier von strukturierten Daten die Rede ist, empfiehlt es sich, die Prinzipien freier Lizenzen auch auf Dokumente anzuwenden. So können Sie z. B. in veröffentlichten Dokumenten entsprechende Creative-Commons-Lizenzen angeben. Der vorliegende Leitfaden beispielsweise steht unter der Creative-Commons-Lizenz CC BY 4.0.



Das Lizenzieren von offenen Daten unter einer freien Lizenz bedeutet nicht automatisch, dass Sie Ihre Angaben im Impressum der gesamten Website, auf der die Daten angeboten werden, ändern müssen. Für die Inhalte der Website können weiterhin abweichende Bestimmungen gelten. Wir empfehlen aber, im Impressum auf die offenen Daten hinzuweisen. Auf creativecommons.org steht ein Lizenzgenerator zur Verfügung.

Schritt 2: Maschinenlesbare Formate (technischer Aspekt)

Das zweite wesentliche Merkmal offener Daten sind technisch zeitgemäße Formate. Datentabellen, die in PDF-Dokumenten oder auf Websites veröffentlicht worden sind, können zwar über Tools ausgelesen werden (über „Web Scraping“¹⁸), was aber einen unnötig hohen Aufwand erfordert und fehleranfällig ist. Sie sollten daher Formate verwenden, die eine möglichst einfache Weiterbearbeitung ermöglichen. Da insbesondere alte Anwendungen mitunter moderne Formate wie JSON, RDF oder XML nicht bereitstellen können, hat sich als Mindeststandard CSV etabliert. Unter Daten in diesem Format (Comma Separated Values) versteht man tabellarische Daten, die in einem Textformat gespeichert und über einen Zeichentrenner (meistens Komma oder Strichpunkt) getrennt hintereinandergeschrieben werden. Tabellenkalkulati-

18 https://de.wikipedia.org/wiki/Screen_Scraping (Download 12.10.2023).

ABBILDUNG 3 5-Sterne-Modell

*	Veröffentlichung von Daten im Internet unter einer freien Lizenz (Format egal)
**	Strukturiertes Format (z. B. Tabelle statt eingescanntes Bild)
***	Offenes, nicht proprietäres Format (z. B. CSV statt Excel)
****	Verwendung von Web-Standards (wie RDF) und URIs, um Daten adressierbar zu machen
*****	Verlinkung eigener Daten zu anderen, adressierbaren Daten (Linked Open Data)

Quelle: Eigene Darstellung. | BertelsmannStiftung
In Anlehnung an 5stardata.info/de.

onsprogramme wie Microsoft Excel oder LibreOffice Calc können solche Formate abspeichern.

Die Formate JSON, RDF oder XML können komplexere Strukturen besser abbilden als CSV und sind daher zu bevorzugen. Man kann einen Datensatz natürlich auch in mehreren Formaten anbieten. Je nach Inhalt können auch gänzlich andere Formate relevant werden. Auch hier gilt das Prinzip: möglichst offene und herstellerunabhängige Formate verwenden.¹⁹

Zur Erläuterung der Bedeutung technischer Formate dient auch das bekannte 5-Sterne-Modell²⁰ (siehe Abbildung 3).

Schritt 3: Metadaten (organisatorischer Aspekt)

Metadaten sind „Daten über Daten“. Ihr Ziel ist es, den eigentlich veröffentlichten Datensatz näher zu beschreiben, um ihn auffindbarer und verständlicher zu machen.

Solche Metadaten sind nicht nur sinnvoll, um sie für Menschen lesbar zu beschreiben, sondern sie sollten im Idealfall auch maschinenlesbar zur Verfügung

¹⁹ Das europäische Datenportal bietet E-Learning-Einheiten zum Thema offene Datenformate an: <https://data.europa.eu/en/training/elearning/open-data-formats> (Download 12.10.2023). Wikipedia führt eine Liste offener Dateiformate: https://en.wikipedia.org/wiki/List_of_open_file_formats (Download 12.10.2023).

²⁰ <https://5stardata.info/de> (Download 12.10.2023).

ABBILDUNG 4 Ein Beispiel für Metadaten

Titel	Offene Unternehmensdaten 2019
Beschreibung	Die Daten aus dem Handelsregister sind zentral für das Vertrauen nicht nur in das Handelssystem, sondern auch in die Demokratie. Trotzdem sind sie von staatlicher Stelle nicht frei zugänglich. Gemeinsam mit OpenCorporates veröffentlichen wir erstmals das Handelsregister als Open Data.
Schlagworte	Handelsregister, Unternehmen
Kategorie	Wirtschaft und Finanzen
Lizenz	CC BY 4.0
Dateiformate	JSON, SQLite
Letzte Änderung	25.1.2019
Herausgeber	OffeneRegister.de
Kontakt	Open Knowledge Foundation Deutschland e.V.
E-Mail	info@okfn.de

Quelle: Eigene Darstellung. | BertelsmannStiftung

stehen. Dazu dient der Standard zur Beschreibung von Metadaten DCAT-AP.de²¹: DCAT-AP.de ist die deutsche Adaption des europäischen Standards DCAT-AP²² zur Beschreibung von Metadaten für Datenportale. Datensätze, deren Metadaten in diesem Standard beschrieben sind, können auch an andere Datenportale weitergereicht werden. Es gibt auch andere Metadatenformate, die verwendet werden können, wie etwa Schema.org²³ oder DublinCore²⁴. Aufgrund der Popularität von Schema.org für die Auffindbarkeit von Internetseiten in Suchmaschinen ist es ohnehin empfehlenswert, die Websites von zivilgesellschaftlichen Organisationen so umzugestalten, dass Schema.org unterstützt wird. Die folgende Tabelle gibt einen Überblick über Metadatenformate:

²¹ Eine umfassende Beschreibung finden Sie unter <https://www.dcat-ap.de> (Download 12.10.2023).

²² <https://joinup.ec.europa.eu/collection/semic-support-centre/solution/dcat-application-profile-data-portals-europe> (Download 12.10.2023).

²³ <https://schema.org/> (Download 12.10.2023).

²⁴ <https://www.dublincore.org/> (Download 12.10.2023).

ABBILDUNG 5 Metadatenformate

	Schema.org	DublinCore (DCMI)	DCAT / DCAT-AP
Zweck	Metadaten für Suchmaschinen	Metadaten für zahlreiche Anwendungen	Metadaten für Datenportale
Version	V 22, 2023-06-29	V 2020-01-20 V3, 2023-03-07 (draft)	V2, 2020-02-04
Standardisierung	W3C Community Group	Dublin Core Metadata Initiative (DCMI)	W3C Recommendation
Anwendungsfall	Websites	Websites und Anwendungen in Bibliotheken, Universitäten	Datenportale
Formate	Microdata, RDFa, JSON-LD	Kein bestimmtes, HTML (meta elements), RDF/XML, JSON, UML	Kein bestimmtes, z. B: RDF, RDF/XML, RDFa, JSON-LD, Turtle
Beispielvokabular	schema:DataCatalog schema:Dataset schema:encodingFormat schema:License	dcmitype:Dataset dcterms:hasFormat dcterms:license	dcat:Catalog dcat:Dataset dct:format dct:license
Link	https://schema.org/version/latest	https://www.dublincore.org/specifications/dublin-core/dcmi-terms/	https://www.w3.org/TR/vocab-dcat-2/
Anmerkungen	Empfehlenswert insbesondere, wenn die Website als Datenportal verwendet wird. Empfehlung	Empfehlenswert, falls Daten aus Anwendungen kommen, die diesen Standard unterstützen.	Empfehlenswert beim Betrieb eines Datenportals. Hat einiges von DublinCore übernommen. Kann über „Application Profiles“ erweitert werden: <ul style="list-style-type: none"> • DCAT-AP für Datenportale in Europa (DCAT-AP.de – Erweiterung für Deutschland) • StatDCAT-AP für Statistikdaten • GeoDCAT-AP für Geodaten • CiteDCAT-AP für Forschungsdaten

Quelle: Eigene Darstellung.

| BertelsmannStiftung

Welche Daten? – Datenkatalog

Wenn Sie die Veröffentlichung von Daten planen und vorbereiten, müssen zunächst interne Zuständigkeiten geklärt werden: Nominieren Sie eine Person, die als **Open-Data-Beauftragte** den Prozess koordiniert. Die Schwerpunkte dieser Koordination liegen in interner und externer **Community-Arbeit** und dem Erstellen und Pflegen eines Datenkatalogs. Die interne Community besteht aus den Hauptamtlichen und Ehrenamtlichen Ihrer Organisation, die externe Community (z. B. Journalist:innen, andere zivilgesellschaftliche Organisationen, Behörden, Entwickler:innen) wird spätestens bei den ersten Veröffentlichungen relevant. Schon von Beginn an ist es empfehlenswert, einen **internen Datenkatalog** zu erstellen, also eine Liste von Datensätzen inklusive weiterer Angaben wie Beschreibung, zuständige Person, verfügbare Formate etc. Dieser Katalog kann Daten verzeichnen, die in anderer Form bereits veröffentlicht worden sind, oder Daten, die neu veröffentlicht werden sollen. Dabei sind drei Aspekte zu berücksichtigen:

A. Selbst-Screening: Was haben wir selbst bereits veröffentlicht? Häufig sind bereits im Internet Dokumente veröffentlicht, die noch nicht den Anforderungen offener Daten genügen. Hier bietet sich das größte Potenzial, diese Daten künftig zusätzlich in einem offenen Format zu veröffentlichen bzw. neue Daten im offenen Format bei der nächsten Veröffentlichung gleich mit anzubieten.

B. Fremd-Screening: Was haben andere bereits veröffentlicht? Hier hilft insbesondere der Blick auf bestehende Datenportale und Websites anderer Organisationen. Ein paar internationale Beispiele:

- The Humanitarian Data Exchange: <https://data.humdata.org/>
- Opendataportal.at: www.opendataportal.at
- Data.world NGO datasets: <https://data.world/datasets/ngo>
- NGO Data Sources: <https://guides.lib.berkeley.edu/c.php?g=496970&p=3401927>

C. Stakeholder Screening: Was wollen die Stakeholder? Welche Daten sind für unterschiedliche Ziel- und Anspruchsgruppen von Interesse? Welche Anfragen nach Daten erhalten wir regelmäßig? Gibt es darunter bereits Anfragen zu offenen Daten? Insbesondere Umfragen und Veranstaltungen bieten sich als Methoden an, um die Interessen der Stakeholder kennenzulernen.²⁵

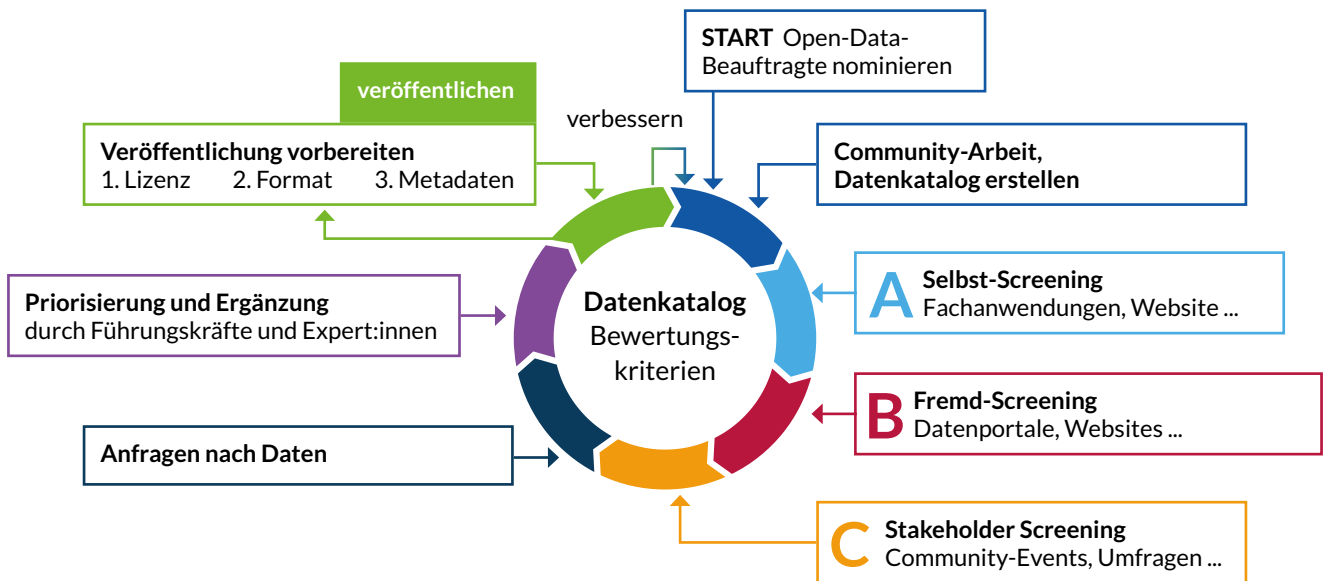
Der so entstehende Datenkatalog liefert die Basis für **Priorisierungen und Ergänzungen durch die Führungsebene und Expert:innen**. Frühere **Anfragen nach Daten** können Ihnen Hinweise liefern, welche Daten Sie veröffentlichen können, um Verwaltungsaufwand zu reduzieren.

Beispielsweise hat sich in der Bertelsmann Stiftung der Aufwand für die Beantwortung von Anfragen nach Daten aus dem Projekt „Wegweiser Kommune“ reduziert, seit diese als offene Daten zur Verfügung stehen.²⁶

²⁵ Siehe auch Wiedemann 2019 bzw. generell die Arbeit der Open Knowledge Foundation Deutschland, z. B. Code for Germany: <https://www.codefor.de> (Download 12.10.2023).

²⁶ Siehe <https://www.wegweiser-kommune.de/open-data> (Download 12.10.2023).

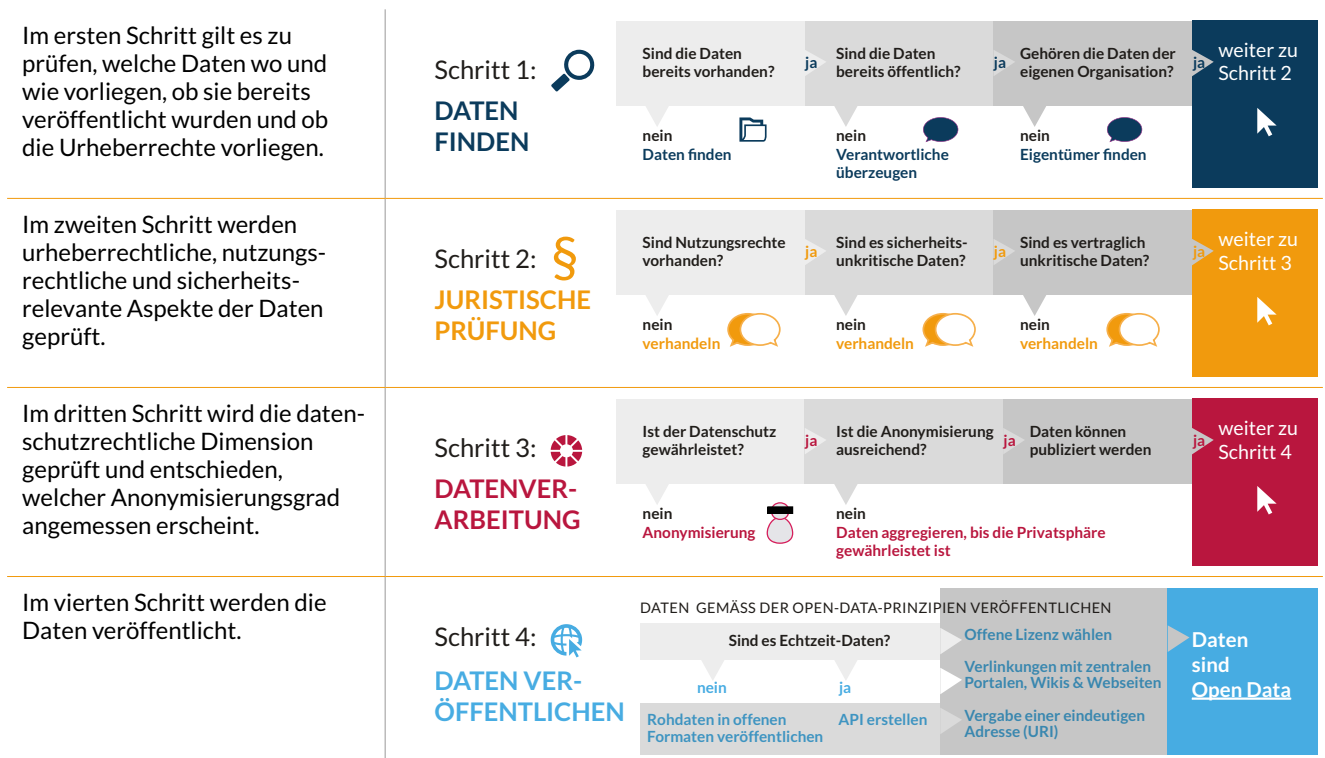
ABBILDUNG 6 Prozess der Erstellung eines Datenkatalogs



Quelle: Eigene Darstellung. In Anlehnung an Krabina und Lutz 2016; Krabina 2020.

BertelsmannStiftung

ABBILDUNG 7 Prozess der Prüfung eines Datensatzes



Quelle: Eigene Darstellung. In Anlehnung an Openall.info und Krabina und Lutz 2016.

BertelsmannStiftung

Sobald mithilfe des internen Datenkatalogs eine ausreichende Entscheidungsgrundlage geschaffen wurde, um erste Veröffentlichungen vornehmen zu können, werden pro Datensatz die Lizenz und die Datenformate festgelegt und die benötigten Metadaten erfasst. Danach kann der Datensatz auf einem Datenportal bzw. auf andere Weise im Internet veröffentlicht werden und der Prozess der Überarbeitung des Datenkatalogs beginnt erneut. Die Bewertungskriterien für das **interne Datenmonitoring** helfen bei der Beurteilung einzelner Datensätze (den Einträgen im Datenkatalog). Für jeden Datensatz kann daher das Prüfschema in Abbildung 7 herangezogen werden.

Das Datenmonitoring kann auch noch detaillierter nach folgenden, etablierten Bewertungskriterien²⁷ durchgeführt werden. Dies ist insbesondere hilfreich, wenn es um die Frage geht, ob Daten, die bisher noch unveröffentlicht sind, veröffentlicht werden können.

Diese Kriterien können als Denkanstoß dienen oder auch mit Punkten, z. B. von 0 bis 5, bewertet werden. Prioritär veröffentlicht wird die Liste der Datensätze mit dem höchsten Punktwert.

Beispiele für Daten

Für Kommunen gibt es in Deutschland mit dem u. a. von der Bertelsmann Stiftung entwickelten Musterdatenkatalog²⁸ einen Katalog an üblicherweise von Kommunen veröffentlichten Datensätzen. Die Stadt Berlin hat eine Liste an „Kerndatensätzen“ veröffentlicht.²⁹ So etwas existiert für die auch wesentlich heterogeneren zivilgesellschaftlichen Organisationen (noch) nicht, doch können beispielsweise folgende Kategorien von Daten für die Veröffentlichung geeignet sein:

- **Standortdaten/Geodaten:** z. B. Standorte von Niederlassungen oder durchgeführten Projekten inkl. Koordinaten und weiterführenden Informationen
- **Leistungsdaten:** z. B. durchgeführte oder geförderte Projekte, angebotene Dienstleistungen

²⁷ Siehe dazu ausführlicher Krabina 2012: 295.

²⁸ <https://www.musterdatenkatalog.de> (Download 12.10.2023).

²⁹ <https://odis-berlin.de/projekte/kerndatensatze/> (Download 12.10.2023).

ABBILDUNG 8 Bewertungskriterien für Daten

Kriterium	Erläuterung
Geheimhaltung/rechtliche Hindernisse	Unterliegen die Daten Geheimhaltungspflichten oder sonstigen rechtlichen Beschränkungen bzw. handelt es sich um infrastrukturkritische Daten?
Personenbezug	Handelt es sich um personenbezogene Daten bzw. lassen sich Rückschlüsse auf Personen aus diesen ableiten?
Nutzungsrecht	Besitzt die Organisation das alleinige Nutzungsrecht der Daten?
Nutzen	Wie hoch wird der Nutzen für die Zielgruppen eingeschätzt?
Aufwand	Wie hoch ist der Aufwand für die Veröffentlichung?
Inhaltliche Datenqualität	Wie hoch wird die Datenqualität eingeschätzt? (Vollständigkeit, zeitliche Nähe, Genauigkeit, Fehlerhaftigkeit ...)
Technische Verfügbarkeit	Sind die Datenformate und Datenquellen in offenen Standards (offene Formate, 5-Sterne-Modell) verfügbar?
Synergie	Werden Daten/Dienste bereits anderweitig von der Organisation angeboten? (Hier gibt es sozusagen Zusatzpunkte, wenn ohnehin geplant ist, eine Veröffentlichung vorzunehmen.)

Quelle: Eigene Darstellung.
In Anlehnung an Krabina 2012.

| BertelsmannStiftung

- **Umfrage-, Mess- und Erhebungsdaten:** Oft werden in Umfragen oder mit anderen Methoden Daten erhoben, die dann ausgewertet werden. Nicht nur die Auswertung, sondern auch die Rohdaten sollten veröffentlicht werden.
- **Finanzdaten, Spendenaufkommen, Förderdaten:** die Offenlegung von Finanzdaten wie z. B. dem Spendenaufkommen oder weiterführenden Daten in Geschäftsberichten. Viele zivilgesellschaftliche Organisationen vergeben auch Projektförderungen. Dafür gibt es die internationalen Standards IATI³⁰ für Entwicklungszusammenarbeit und 360Giving³¹ in Großbritannien für Daten zu Förderungen.

³⁰ <https://iatistandard.org> (Download 12.10.2023).

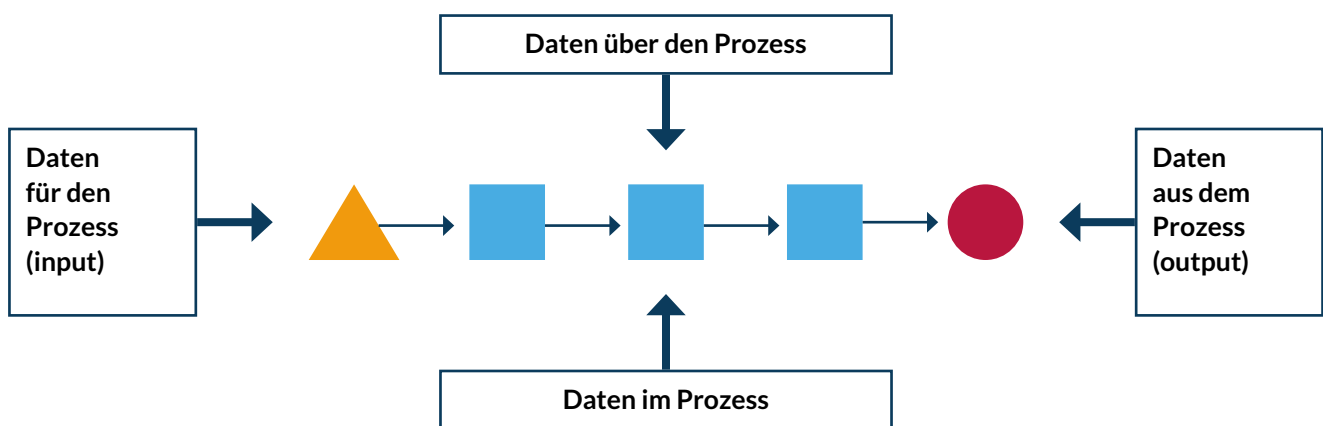
³¹ <https://www.threesixtygiving.org/data-standard/> (Download 12.10.2023).

- **Statistikdaten:** Wenn die Veröffentlichung von Rohdaten aus Datenschutzgründen nicht möglich ist, können detaillierte Statistikdaten veröffentlicht werden.
- **Bilder, Audio- und Videoaufnahmen.**

Ein prozess- bzw. projektorientierter Zugang (siehe Abbildung 9) kann ebenso hilfreich sein, sich über das Vorkommen von Daten bewusst zu werden. Ein

Prozess kann z. B. die Förderabwicklung sein. Es gibt dafür Daten, die zu Beginn des Prozesses relevant sind, Daten, die während der Prozessschritte hinzukommen und Daten, die aus dem Ende des Prozesses resultieren. Auch Daten über den Prozess können relevant sein, z. B. die Anzahl der Prozesse, die pro Jahr ablaufen. In projektgetriebenen Organisationen kann man in der Grafik „Prozess“ auch durch „Projekt“ ersetzen.

ABBILDUNG 9 Daten in Prozessen und Projekten



Quelle: Eigene Darstellung. In Anlehnung an Wissensmanagement in Prozessen/Projekten. KM-A, Knowledge Management Academy.

| BertelsmannStiftung

Wo veröffentlichen? – Dateninfrastrukturen

Für die Veröffentlichung ist das wichtigste Kriterium, dass der Datensatz über eine URL möglichst dauerhaft abrufbar ist. Der Datensatz aus unserem Beispiel (Offene Unternehmensdaten 2019; Abbildung 4) ist unter einer eigenen URL abrufbar, die direkt auf die downloadbare Datei verweist.³²

Wohin mit den Daten?

Die erste Frage ist daher: Wohin sollen die Daten,³³ damit sie für andere abrufbar sind? Hier gibt es je nach den Voraussetzungen Ihrer Organisation verschiedene Möglichkeiten:

- Upload der Daten in eine **gehostete Filesharing-Lösung**. Hier gibt es eine Vielzahl an Optionen, wie etwa GitLab/GitHub oder NextCloud für beliebige Daten. Auch kommerzielle Anbieter bieten meistens Hosting einer gewissen Menge an Daten kostenlos an. Für sehr große, statische Dateien könnte auch das BitTorrent-Protokoll³⁴ verwendet werden.
- Upload der wissenschaftlichen Daten in ein **themenspezifisches Repositorium** wie Wikidata, Wikimedia Commons, OpenStreetMap etc.
- Upload der Daten auf **Zenodo**, einer kostenfreien wissenschaftliche Datenrepositorium-Plattform, die von CERN (Europäische Organisation für Kernforschung) entwickelt wurde.

32 Die URL lautet https://daten.offeneregister.de/de_companies_ocdata.jsonl.bz2 (Download 12.10.2023).

33 Im Falle einer API laden Sie keine Daten ins Internet, sondern bieten eine Schnittstelle zu einem über das Internet erreichbaren Server an, der die API bereitstellt.

34 Siehe <https://de.wikipedia.org/wiki/BitTorrent> (Download 12.10.2023).

- Upload der Daten auf die **eigene Website**. So, wie Sie im Content-Management-System ein Bild oder PDF-Dokument hochladen, können Sie das auch mit anderen Dateien tun. Auch das Hochladen in ein Verzeichnis eines FTP-Servers ist eine kostengünstige Möglichkeit.
- Errichten eines **eigenen Datenportals**. Spezielle Datenportalsoftware ist in kommerziellen und Open-Source-Varianten erhältlich.³⁵ Sie ermöglichen neben dem Verzeichnen und Präsentieren des Datenangebots auch das direkte Hochladen von Dateien.

Wo verzeichne ich mein Datenangebot?

Die zweite Frage ist nun: Wo wird mein Datenangebot präsentiert?

Portale wie Govdata.de oder das Europäische Datenportal sind sogenannte **Metadatenverzeichnisse**. Diese verwenden Software für Datenportale, ohne die Daten selbst zu hosten. Metadatenverzeichnisse haben den Zweck, Datensätze besser auffindbar zu machen. Bei GovData.de kann man also im Datenangebot aller Behörden Deutschlands suchen, ohne die jeweiligen Websites aufrufen zu müssen. Die Daten werden dort aber in der Regel nicht hochgeladen, sondern es werden die URLs zu den einzelnen Dateien verzeichnet, die dann auf die originalen Quellen zeigen. Aktuell gibt es jedoch noch kein deutschlandweites Verzeichnis für offene Daten aus der Zivilgesellschaft.³⁶

35 Die bekannteste ist „CKAN“, mit der auch GovData.de betrieben wird: <https://ckan.org> (Download 12.10.2023).

36 In Österreich gibt es so ein Portal: <https://www.opendataportal.at> (Download 12.10.2023).

Der Betrieb eines eigenen Datenportals hat neben dem Hochladen der Daten an einer zentralen Stelle auch den Vorteil, dass die Daten auch gleich mit Metadaten versehen werden können und das so entstehende Datenangebot übersichtlich präsentiert wird. In Zukunft können die Metadaten dann an andere Metadatenverzeichnisse weitergegeben werden. Auf der eigenen Website kann das Datenportal beworben werden. Für das eigene Datenportal wird ein Webserver benötigt, auf dem die Datenportalsoftware läuft.³⁷

Themenspezifische Repositorien haben den Vorteil, dass diese die Daten nicht nur hosten, sondern auch gleich verzeichnen und zur Nutzung anbieten. Allerdings ist es häufig nicht möglich, ein eigenes Datenangebot zu präsentieren. Daher sollte man auch in diesem Fall das Datenangebot auf der eigenen Website bewerben.

Falls die Daten von allgemeinem Interesse sind, können diese an **Wikidata** „gespendet“ werden³⁸ oder im Fall von Fotos, Bildern, Video- oder Audiodateien diese auf **Wikimedia Commons**³⁹ veröffentlicht werden. Auch **Wikisource** oder **Wikibooks** können mögliche Projekte sein, an denen man sich beteiligen kann.⁴⁰

Auf **OpenStreetMap** wird freies Kartenmaterial erstellt,⁴¹ bei **Open Food Facts** werden Daten über Lebensmittel gesammelt. **Sensor.community** ist ein von vielen Mitwirkenden betriebenes, globales Sensornetzwerk, das offene Umweltdaten generiert.⁴² Die **Open Repair Alliance** definiert mit dem Open Repair Data Standard (ORDS)⁴³ einen gemeinsamen Ansatz für die Sammlung und den Austausch von Reparaturdaten über kleine Elektrogeräte und Elektronik.

37 Weiterführendes siehe Ergebnisbericht Projekt „Data Go - Open Data zu GovData“, Version: 1.0 vom 12.12.2022: https://www.it-planungsrat.de/fileadmin/beschluesse/2023/Beschluss2023-12-GovData_Bericht.pdf (Download 12.10.2023).

38 https://www.wikidata.org/wiki/Wikidata:Data_donation (Download 12.10.2023).

39 https://commons.wikimedia.org/wiki/Commons:Trage_deine_eigene_Arbeit_bei?uselang=de (Download 12.10.2023).

40 https://wikisource.org/wiki/Wikisource_and_Wikibooks (Download 12.10.2023).

41 <https://welcome.openstreetmap.org/how-to-give-back/> (Download 12.10.2023).

42 <https://sensor.community/de/> (Download 12.10.2023).

43 <https://openrepair.org/open-data/open-standard/> (Download 12.10.2023).

Darüber hinaus ist empfehlenswert, die Entwicklung der sogenannten „**Data Spaces**“ und „**Knowledge Graphs**“ zu beobachten. Hier entstehen (meist thematisch fokussierte) Datenräume oder Wissensgraphen. Beispiele sind

- Open Data im Tourismus: <https://open-data-germany.org>
- Datenraum Kultur: <https://www.acatech.de/projekt/datenraum-kultur/>
- Mobility Data Space: <https://www.mobility-data-space.de>
- European Health Data Space: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_de
- International Data Spaces: www.dataspaces.fraunhofer.de
- Gaia-X Hub Germany: <https://gaia-x-hub.de/>

Ein eigenes Datenportal ist somit für eine Veröffentlichung nicht unbedingt nötig. Sie können Datensätze auch auf Ihrer eigenen Website zum Download anbieten, dort die entsprechenden Metadaten angeben und eine Lizenz festlegen. Bei Content-Management-Systemen, mit denen Websites betrieben werden (z. B. Wordpress, Drupal, Typo3) gibt es möglicherweise Erweiterungen, um Metadaten angeben zu können.⁴⁴ Auch Wikis können genutzt werden.⁴⁵

44 Für Drupal gibt es beispielsweise dkan: <https://dkan.readthedocs.io> (Download 12.10.2023).

45 Das OGD-Cockpit basiert auf einer Konfiguration der Software Semantic MediaWiki: <https://www.ogdcockpit.eu> (Download 12.10.2023).

Wie beginnen? – Phasen der Umsetzung

Wer keine offenen Daten veröffentlicht, stellt fallweise Informationen auf die Website, kümmert sich weder um Lizenzen noch um Formate und erschwert damit die Wiederverwendung. Die Potenziale offener Daten bleiben daher ungenutzt. Auf Anfrage würden Sie in diesem Fall auf PDF-Dokumente auf Ihrer Website verweisen oder hin und wieder ein Excel-File per E-Mail verschicken.

In der **Einstiegsphase** versuchen Sie erstmalig offene Daten zu publizieren. Ihr Fokus liegt daher auf Daten, die in anderer Form bereits veröffentlicht worden sind, oder Sie beginnen damit, proaktiv Wünsche der Stakeholder zu berücksichtigen. Die ersten Veröffentlichungen können manuell erfolgen (sowohl Daten wie auch Metadaten werden einzeln hochgeladen). Datenbewertungen nehmen Sie, wenn, dann nur in fallweiser Diskussion vor. Sie orientieren sich insbesondere an vergleichbaren Organisationen und vernetzen sich mit diesen. Sie erwarten sich, erste Aufmerksamkeit für das Thema zu generieren (intern wie extern). Dabei helfen Ihnen erste prototypische Datennutzungen.

In der **fortgeschrittenen Organisation** überzeugen Sie als Open-Data-Beauftragte:r immer mehr wichtige Akteur:innen in Ihrer Organisation, die zuvor eventuell noch skeptisch waren. Sie müssen Ihre Veröffentlichungsprozesse optimieren, da der Ansatz der manuellen Veröffentlichung bald nicht mehr ausreichend ist. Sie schaffen es, dass die anderen internen Akteur:innen regelmäßig neue Datensätze erzeugen (z. B. durch Exporte aus Anwendungen) und Metadaten einpflegen. Um den Überblick zu bewahren und Veröffentlichungsphasen planen zu können, führen Sie nun die Datenbewertungen strukturierter durch. In anderen Datenkatalogen blicken Sie immer mehr

auf die Potenziale (also auf Datensätze, die andere schon veröffentlicht haben, Sie aber noch nicht).

Erste Apps und Anwendungen entstehen, die Ihre Initiative voranbringen.

Bei **Open-Data-Profis** hat sich ein Open-by-Default-Ansatz durchgesetzt. IT-Systeme werden mit offenen APIs für Daten und Metadaten ausgestattet. Sie orientieren und vernetzen sich international und sind mit anderen zivilgesellschaftlichen Organisationen in laufendem Kontakt. Datenökosysteme entstehen, die ihre Organisation dabei unterstützen, ihre Wirkung zu entfalten. Intern achten Sie auf Data Governance, in die alle Abteilungen einbezogen sind. Viele interne Abläufe haben sich daher verändert und verbessert.

Die beschriebenen Phasen sind hier nur beispielhaft skizziert und in dieser Form wohl kaum genauso in der Praxis anzutreffen. Abbildung 10 ist insbesondere dazu gedacht, Orientierung und Überblick zu bieten, welche Aspekte der eigenen Open-Data-Initiative noch verbessert werden könnten.

Die Übersicht hilft Ihnen bei der Orientierung im Laufe der Phasen, die Sie im Zuge der Veröffentlichung offener Daten voraussichtlich durchlaufen werden. Lassen Sie sich nicht abschrecken durch die hoch gesteckten Ziele in „Phase III: Profi“. Wenn Sie mit Ihrer Organisation in Phase I eintreten, sind Sie bereits den wichtigsten Schritt gegangen!

ABBILDUNG 10 Open-Data-Phasen

	Kein Open Data	Phase I: Einstieg	Phase II: Fortgeschrittene	Phase III: Profi
Datenfokus	keiner	veröffentlichte	unveröffentlichte	Open by Default
Prozess	Website	ad hoc	etabliert	evaluiert
Involvierte Organisationseinheiten	nach Bedarf	willige	strategisch wichtige	alle
Einflussbereich auf die Organisation	–	punktuell	regelmäßig	strategisch
Einbeziehung der Community	reaktiv (Reaktion auf Anfragen)	proaktiv: Datenwünsche, Umfragen	gestaltend: Hackdays, Events	kontinuierlich
Nutzungszahlen	–	erfassen	analysieren und veröffentlichen	Datenangebot anpassen
Datenformate	PDF/Excel	CSV	XML/JSON	RDF
Geobezug	–	CSV, Koordinaten	GeoJSON, Rasterdaten	WMS, Vektordaten, OGC-Schnittstelle
Veröffentlichung	händisch	händisch	Export (aus Fachanwendungen)	API
Metadatenerfassung	keine	händisch (per Formular in das Datenportal)	Import (in das Datenportal)	API (harvesting)
Datenbewertung (laut Kriterien)	keine	fallweise (persönlich, E-Mail)	strukturiert (Tabelle)	kontinuierlich (Kollaborationstool)
Datenkatalog (Umfang und Fokus)	–	Topthemen (Was haben fast alle? Bspw. Umfragen, Finanzen ...)	Potenziale (Was haben andere, was bei uns noch fehlt?)	komplett
Blick auf andere	egal	regional	national	international
Erwartungen (extern)	Information (keine Wiederverwendung)	Aufmerksamkeit, Prototypen	Apps, Anwendungen	Wirkungen, Vernetzung, Daten-ökosysteme
Erwartungen (intern)	–	Awareness	Nutzen	Veränderung
Zielebene	–	Daten veröffentlichen	Veröffentlichungsprozesse optimieren	Data Governance
Open-Government-Vorgehensmodell⁴⁶	–	1 Transparenz	2 Partizipation und 3 Kollaboration	4 Nachhaltigkeit und Public Value
Open Data Maturity Level⁴⁷	–	1 Initial -> 2 Repeatable	3 Defined -> 4 Managed	4 Managed -> 5 Optimizing
5-Sterne-Modell⁴⁸	* bis **	***	****	*****

Quelle: Eigene Darstellung. In Anlehnung an Krabina 2020.

| BertelsmannStiftung









46 Phasen nach Krabina und Lutz 2016.

47 <https://theodi.org/article/open-data-maturity-model-2> (Download 12.10.2023).48 <https://5stardata.info/de> (Download 12.10.2023).

Jetzt starten – eine Checkliste

Sie sind nun bereit für den Start! Folgende Checkliste hilft dabei, die wichtigsten Aspekte bei der Umsetzung in der eigenen Organisation im Blick zu haben:

ABBILDUNG 11 Checkliste

	Thematisieren	Machen Sie die Bedeutung von Daten für Ihre Organisation zum Thema, überzeugen Sie von den Chancen offener Daten und räumen Sie Bedenken aus. Details siehe Kapitel „Worauf gefasst sein? – Häufige Fragen und Antworten“.
	Planen	Formen Sie eine Koalition der Willigen, stellen Sie interne und externe Ressourcen sicher. Erstellen Sie einen Umsetzungsplan und holen Sie sich Unterstützung der Führungskräfte für die Umsetzung. Setzen Sie sich Ziele!
	Team einrichten	Benennen Sie Verantwortliche. Ein Kernteam hilft den Organisationseinheiten bei den Datenveröffentlichungen. In den Organisationseinheiten können Ansprechpersonen für Daten (auch: „Data Stewards“) etabliert werden.
	Datenkatalog erstellen	Erstellen Sie einen ersten Datenkatalog! Führen Sie ein internes Datenmonitoring durch und bewerten Sie einzelne Datensätze. Legen Sie Lizenzen, mögliche technische Formate und erste Metadaten fest. Details siehe Kapitel „Welche Daten? – Datenkatalog“.
	Dateninfrastruktur klären	Besprechen Sie mit der IT die Möglichkeiten der Dateninfrastruktur. Details siehe Kapitel „Wo veröffentlichen? – Dateninfrastrukturen“.
	Datenfreigaben planen	Erstellen Sie einen Veröffentlichungsplan. Dafür benötigen Sie erste Datensätze, dazugehörige Metadaten und die Infrastruktur, in die veröffentlicht wird. Betreiben Sie begleitende Öffentlichkeits- und Communityarbeit.
	Veröffentlichen	Veröffentlichen Sie erste Datensätze. Gratulation, sie haben es geschafft!
	Dran bleiben	Jetzt beginnt die weitere, laufende Arbeit am Datenkatalog. Diesem sollten neue Datensätze hinzugefügt werden und bestehende müssen regelmäßig aktualisiert werden. Auch die interne Arbeit ist nicht vorbei: Stärken Sie die Datenkompetenz Ihrer Organisation, überzeugen Sie weitere Organisationseinheiten und beschäftigen Sie sich mit Data Governance ⁴⁹ . Überprüfen und adaptieren Sie Ihre Ziele!

Quelle. Eigene Darstellung.

| BertelsmannStiftung

49 Siehe z. B. <https://www.computerwoche.de/a/der-richtige-weg-zur-data-governance,3546149> (Download 12.10.2023).

Worauf gefasst sein? – Häufige Fragen und Antworten

Bedarf es einer gesetzlichen Grundlage für Open Data?

Grundsätzlich nein. Genauso wie eine Organisation selbst entscheidet, welche Informationen sie auf ihrer Website veröffentlicht, kann sie ebenso selbst entscheiden, Daten zu veröffentlichen. Dazu wird kein neuer Rechtsrahmen benötigt. Rechtsbereiche wie Datenschutz und Urheberrecht sind selbstverständlich zu beachten. (Siehe Kapitel „Wie vorgehen? – Schritte zu offenen Daten“.)

Widerspricht Open Data nicht dem Datenschutz?

Nein. Personenbezogene Daten sind per Definition kein Gegenstand offener Daten. Es handelt sich in der Regel also um Daten, die keinerlei Personenbezug aufweisen (siehe oben, „Kriterien für das Datenmonitoring“) und deren Verarbeitung und Veröffentlichung nicht unter die Datenschutz-Grundverordnung (DSGVO) fallen. Allerdings ist dies eine zu kurz greifende Betrachtung, da sehr wohl auch Daten, in denen Personenbezug vorkommt oder hergestellt werden kann, von Nutzen sein können. Daher müssen bei Daten mit Personenbezug Methoden der Anonymisierung oder Synthetisierung angewandt werden, bevor diese veröffentlicht werden können.⁵⁰ Alternativ können auch Datenräume („Data Spaces“) genutzt werden, über die auch sensible Daten ausgetauscht werden können. (Siehe Kapitel „Welche Daten? – Datenkatalog“.)

50 Siehe dazu Drechsler und Jentsch 2018.

Wenn wir die Daten völlig freigeben, verlieren wir dann nicht die Kontrolle?

Ja, aber in vielen Bereichen haben wir die Kontrolle ohnehin bereits verloren! Viele Daten werden über kommerzielle Anbieter wie Google (Google Maps) oder auch freie Community-Projekte wie Wikidata oder OpenStreetMap⁵¹ gesammelt oder auch über Scraping aus PDF-Dokumenten oder Websites herausgelesen. Häufig steht beim Sammeln oder Erzeugen von Daten die Gemeinwohlorientierung nicht im Vordergrund. Nur wenn Daten für alle frei zur Verfügung stehen, können diese von allen genutzt werden, um gesellschaftlichen Mehrwert zu erzielen. (Siehe Kapitel „Warum offene Daten? – Open Commons“.)

Sollten wir Daten nicht lieber verkaufen?

Besser nicht. Es ist zwar auch denkbar, einen Basisdatensatz als offenen und kostenlosen Datensatz anzubieten und einen umfangreicheren gegen Bezahlung, aber es sprechen mehrere Argumente dagegen, Daten zu verkaufen. Es kommt zu Einnahmensillusionen (der Aufwand für den Datenverkauf wird unterschätzt) und Kund:innen wechseln zu Alternativprodukten. In den seltensten Fällen wird der Zweck der zivilgesellschaftlichen Organisation die Erhebung und der Verkauf von Daten sein. Daher ist es eher eine zeitgemäße Art der Gemeinwohlorientierung, Daten zur Weiternutzung anzubieten.⁵² (Siehe Kapitel „Warum offene Daten? – Open Commons“.)

51 Siehe <https://www.wikidata.org> bzw. <https://www.openstreetmap.org> (Download 12.10.2023).

52 Siehe dazu ausführlicher Jörg 2014.

Aber wir haben das doch schon als PDF veröffentlicht, wozu der doppelte Aufwand?

Ein doppelter Aufwand ist nur ein Artefakt des Einstiegs in das Thema. Sobald sich Prozesse etabliert haben, wo Daten aus internen Anwendungen automatisch veröffentlicht werden, wird der Aufwand wieder sinken. Und überlegen Sie einmal, wie viel Aufwand die verschiedenen Organisationseinheiten in die Herstellung von gedruckten Publikationen stecken, wie etwa ein Geschäfts- oder Projektbericht, wo der Nutzen „nur“ in der Information der interessierten Öffentlichkeit besteht – ohne den Zusatznutzen der Wiederverwendbarkeit.

Sehen Sie doch das Anbieten offener Daten als eine Form der Barrierefreiheit: Daten sind auf Websites und in PDF-Dokumenten „eingesperrt“ und werden durch die Öffnung barrierefrei angeboten. (Siehe Kapitel „Wie vorgehen? – Schritte zu offenen Daten“.)

Wir haben doch gar keine (relevanten) Daten?

Doch. Es gibt unterschiedliche Nutzungsmöglichkeiten für verschiedene Arten von Daten. Auch Daten, die für einen selbst nicht besonders relevant erscheinen, können möglicherweise für andere bzw. in Kombination mit anderen Daten in einem gesamtgesellschaftlichen Kontext Bedeutung erlangen.

Nicht immer führen Datenveröffentlichungen zu bahnbrechenden neuen Anwendungen. Aber die Information der Öffentlichkeit ist ja stets auch eine relevante und wichtige Aufgabe. Möglicherweise fehlt es auch an einer Nachfrageorientierung: Welche Daten stoßen auch wirklich auf Interesse und Nachfrage? Dabei helfen Hackathons und andere Veranstaltungen zur Einbeziehung verschiedener Stakeholder.⁵³ (Siehe Kapitel „Welche Daten? – Datenkatalog“.)

Nur Daten rauszustellen, bringt doch nichts. Wie kommen wir zu Apps?

Dazu ist es ratsam, einen kontinuierlichen Austausch mit allen Stakeholdern zu pflegen. Je früher Sie auf andere hören, welche Daten diese für relevant halten, desto rascher werden Ihre Daten auch genutzt werden und damit möglicherweise auch Apps oder Anwendungen entstehen. (Siehe Kapitel „Wie beginnen? – Phasen der Umsetzung“.)

Müssen wir ein eigenes Datenportal betreiben?

Nicht unbedingt. Es hängt insbesondere von der erwarteten Menge an Daten und den organisatorischen Ressourcen ab, ob sich der Aufwand dafür lohnt. (Siehe Kapitel „Wo veröffentlichen? – Dateninfrastrukturen“.)

Wir haben gar kein Recht an diesen Daten!

Versuchen Sie, fehlende Rechte zu klären. Häufig wird vergessen, solche Fragen schon im Vorfeld auszuräumen. Denken Sie daher bei der nächsten Ausschreibung eines Projekts (z. B. Studie, Publikation oder Website) oder einer Softwarelösung daran, dass offene Daten zur Verfügung gestellt werden sollen. Das spart Ihnen Arbeit und dient gleichzeitig zur Klärung des rechtlichen Aspekts. (Siehe Kapitel „Welche Daten? – Datenkatalog“.)

⁵³ Siehe auch Wiedemann 2019 bzw. generell die Arbeit von Open Knowledge Deutschland, z. B. „Code for Germany“: <https://www.codefor.de> (Download 12.10.2023).

Welche Begriffe sollten Sie kennen? – Glossar

API Application Programming Interface. Schnittstelle zur dynamischen Bereitstellung von Daten.

CC0 Creative-Commons-Lizenz. Zero (0) bedeutet „keine Rechte vorbehalten“.

CC-BY Creative-Commons-Lizenz. BY bedeutet Namensnennung.

CSV Comma-Separated Value. Mit Trennzeichen getrennte, tabellarische Daten.

DCAT-AP Format zur Beschreibung von Metadaten.

FAIR Initiative für FAIRe Forschungsdaten: Findable (Auffindbar), Accessible (Zugänglich), Interoperable (Interoperabel) und Reusable (Wiederverwendbar).

JSON JavaScript Object Notation. Datenformat in einer besonders gut maschinenlesbaren Textform.

LOD Linked Open Data. Vernetzte offene Daten.

Metadaten Daten, die Datensätze beschreiben.

OGD Open Government Data. Offene Verwaltungsdaten.

RDF Resource Description Framework. Datenformat zur Beschreibung von Ressourcen im Internet.

URL Uniform Resource Locator. Lokalisiert eine Ressource, beispielsweise eine Webseite.

Weiterführendes Glossar siehe Bundesverwaltungsamt⁵⁴ und Open Data Handbook⁵⁵.

54 https://www.bva.bund.de/SharedDocs/Downloads/DE/Behörden/Beratung/Methoden/open_data_glossar.pdf (Download 12.10.2023).

55 <https://opendatahandbook.org/glossary/en/> (Download 12.10.2023).

Wo können Sie weiterlesen? – Weiterführende Informationen

Leitfäden und Studien

- Ergebnisbericht Projekt „Data Go - Open Data zu GovData“, Version: 1.0 https://www.it-planungsrat.de/fileadmin/beschluesse/2023/Beschluss2023-12_GovData_Bericht.pdf (2023)
- Open Data Handbook <https://opendatahandbook.org> (2016–2022)
- The Economic Impact of Open Data. European Data Portal. <https://data.europa.eu/sites/default/files/the-economic-impact-of-open-data.pdf> (2022) sowie weitere Studien <https://data.europa.eu/en/publications/studies>
- Land Berlin: Das Berliner Open-Data-Handbuch <https://berlinonline.github.io/open-data-handbuch/> (2019–2023)
- Bundesverwaltungsamt: Open Data – Handbuch für offene Verwaltungsdaten https://www.bva.bund.de/DE/Services/Behoerden/Beratung/Beratungszentrum/Methoden/_documents/stda_open_data.html (2022)
- Bitkom: Data Spaces and Data Ecosystems – First Explainer and Current Status. https://www.bitkom.org/sites/main/files/2022-09/Bitkom_Data_Spaces_Data_Ecosystems%20Sep_2022.pdf (2022)
- Bitkom: Leitfaden Open API – offene Daten professionell anbieten und nutzen <https://www.bitkom.org/Bitkom/Publikationen/Open-API-offene-Daten-professionell-anbieten-und-nutzen> (2019)
- Fraunhofer Fokus: Open Data zwischen Wunsch und Wirklichkeit (2019) <https://doi.org/10.24406/publica-fhg-299947>
- Fraunhofer Fokus: Leitfaden für hochwertige Daten und Metadaten (2019) https://www.fokus.fraunhofer.de/de/fokus/news/leitfaden_nqdm_2019
- Stiftung Neue Verantwortung. Wettbewerb um Daten. Über Datenpools zu Innovationen <https://www.stiftung-nv.de/de/publikation/wettbewerb-um-daten-ueber-datenpools-zu-innovationen> (2019)

Internetlinks

- Openall.info: <https://openall.info>
- 5-Sterne-Modell: <https://5stardata.info/de/>
- Code for Germany: <https://www.codefor.de>
- DCAT-AP.de: <https://www.dcat-ap.de>
- Europäisches Datenportal:
www.europeandataportal.eu
- GovData (Datenportal für Deutschland):
<https://www.govdata.de>
- Open Government Partnership Deutschland:
<https://opengovpartnership.de>
- Forschungsdaten.info
<https://forschungsdaten.info>
- GO FAIR Initiative: <https://www.go-fair.org>
- Digital Public Goods:
<https://digitalpublicgoods.net>

Literaturverzeichnis

- Drechsler, Jörg, und Nicola Jentzsch (2018). „Synthetische Daten. Innovationspotential und gesellschaftliche Herausforderungen“. Hrsg. Stiftung Neue Verantwortung. Berlin. https://www.stiftung-nv.de/sites/default/files/synthetische_daten.pdf (Download 12.10.2023).
- Jörg, Wolfgang (2014). „ViennaGIS® verschenkt seine Geodaten – Können wir uns das leisten?“. Vermessung & Geoinformation 3. 138–145. <https://www.ovg.at/de/vgi/files/pdf/5180> (Download: 12.10.2023).
- Juech, Claudia (2022). „Gesellschaftlicher Mehrwert von Open Data aus der Zivilgesellschaft“. https://www.bertelsmann-stiftung.de/fileadmin/files/Juech_Impuls_Mehr_Open_Data_aus_der_Zivilgesellschaft_221108.pdf (Download: 12.10.2023).
- Krabina, Bernhard, und Brigitte Lutz (2016). „Open Government Vorgehensmodell“. <https://www.kdz.eu/de/wissen/studien/open-government-vorgehensmodell> (Download 12.10.2023).
- Krabina, Bernhard (2023). „Ein Leitfaden für offene Daten“. Data.gv.at. Wien <https://www.data.gv.at/leitfaden> (Download 12.10.2023).
- Krabina, Bernhard (2020). „Ein Leitfaden für offene Daten“. Bertelsmann Stiftung. Gütersloh. <https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/ein-leitfaden-fuer-offene-daten> (Download 12.10.2023).
- Krabina, Bernhard (2012). „Internal data monitoring for Open Government“. CeDEM 12. Proceedings of the International Conference for E-Democracy and Open Government. Hrsg. Peter Parycek, Noella Edelmann und Michael Sachs. Krems: Edition Donau-Universität Krems. 295.
- Wiedemann, Mario (2019). „Welchen Vorteil bieten offene Daten gemeinnützigen Organisationen?“. <https://www.bertelsmann-stiftung.de/de/unsere-projekte/smart-country/projektnachrichten/open-data-workshop/> (Download 12.10.2023).

Über den Autor

Bernhard Krabina ist Experte für Daten-, Informations- und Wissensmanagement der KM-A Knowledge Management Associates in Wien.

Er hat zahlreiche Behörden im deutschsprachigen Raum bei ihren Open-Data-Initiativen begleitet und ist im Team Open Commons für den Datenauftritt der Stadt Linz in Österreich zuständig. Darüber hinaus forscht er an der WU Wien am Institut für Daten-, Prozess-, und Wissensmanagement und ist Lektor an der FH Campus Wien.

Danke an

Damian Paderta, <https://openall.info>

Dana Milovanovic, Deutsche Stiftung für Engagement und Ehrenamt, <https://d-s-e-e.de>

Dr. Georg Förster, Deutscher Caritasverband e. V., <https://www.caritas.de>

Adresse | Kontakt

Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256
33311 Gütersloh
Telefon +49 5241 81-0
bertelsmann-stiftung.de

Mario Wiedemann
Senior Project Manager
Daten für die Gesellschaft
Telefon +49 5241 81-81305
mario.wiedemann@bertelsmann-stiftung.de