

# Handreichung

Dateninventur, Datenkatalog und Open Data.

Eine Handreichung der byte für die Durchführung einer ersten Dateninventur und die Erstellung eines Datenkatalogs zur Erfassung von Verwaltungsdaten. Verschaffen Sie sich ganz pragmatisch einen Überblick über Ihre Verwaltungsdaten, deren Struktur und Nutzungspotential und machen Sie so den ersten Schritt zu einer transparenten Verwaltung, die ihre Datensätze als Ressourcen begreift und als Open Data teilt.

Stand Februar 20.02.2024

# Von der Dateninventur zu Open Data

## Verwaltungsdaten im Überblick

Kennen Sie das das Open-Data-Portal [open bydata](#) der byte – der Bayerischen Agentur für Digitales? Ganz nach dem Motto „Wir digitalisieren Bayern“ werden auf dieser Plattform seit Mai 2023 Verwaltungsdaten aus Bayern einfach, zentral und kostenfrei als Open Data veröffentlicht.

Mit dieser Handreichung wollen wir als Open-Data-Team der byte absichtlich einen ganz pragmatischen Start mit einfachen Mitteln in eine **Dateninventur** geben. Gewinnen Sie einen **Überblick über Ihre Verwaltungsdaten** und erfassen Sie diese strukturiert als **Datenkatalog**. Hier finden Sie praktische Tipps und eine Excel-Tabelle zur individuellen Nutzung. Nutzen Sie diese Hilfestellung für einen leichten Einstieg, die Datenschätze Ihrer Organisation zu heben und bereiten Sie deren **Veröffentlichung als Open Data** vor.

## Ihr Ansprechpartner

Bei Rückfragen steht Ihnen das Open-Data-Team der byte jederzeit gerne zur Verfügung:

Schreiben Sie uns an und vereinbaren Sie einen kostenfreien Beratungstermin, wenn Sie planen Daten aus Ihrer Einrichtung zu veröffentlichen: [openbydata@byte.bayern](mailto:openbydata@byte.bayern)

### Open Data

Im Kontext der öffentlichen Verwaltung handelt es sich bei „offenen Daten“ um Informationen, die von öffentlichen Stellen gesammelt, erstellt oder bezahlt wurden und für eine allgemeine Nachnutzung frei verfügbar sind. D.h. diese Verwaltungsdaten können als „offene Daten“ veröffentlicht werden, sofern Gesetze, Datenschutz oder Sicherheitsbedenken dies nicht ausschließen. Mit offenen Daten können Unternehmen, Zivilgesellschaft oder Forschung, Mehrwerte schaffen und neue Services entwickeln.

## Wozu eine Dateninventur?

Eine über alle Fachabteilungen hinweg durchgeführte Dateninventur bedeutet **Transparenz** über vorhandene Daten, deren Entstehung und Vorhaltung. Sie hilft Ihnen, Informationen zu bündeln und Fragestellungen effizienter zu beantworten. Werden diese Daten zusätzlich frei verfügbar gemacht, können auch Bürger und Bürgerinnen sowie die Wirtschaft von besseren Angeboten profitieren.

Ein Beispiel: Eine Kommune möchte Familien besser unterstützen und daher Buslinien und Haltestellen so umplanen, dass wichtige Orte wie Spielplätze einfacher zu erreichen sind und mit zukünftigen Einwohnerzahlen und Sharing Mobility Angeboten korrespondieren. Mit den Ergebnissen einer Dateninventur lässt sich schnell ermitteln, welche Stellen für dieses Vorhaben wichtige Daten liefern können. Die Dateninventur erhöht die Transparenz, fördert effiziente Kommunikation zwischen Organisationseinheiten und bildet eine Basis für weitere Digitalisierungsschritte und/oder Automatisierungen.

## Was ist eine Dateninventur?

Aus dem täglichen Leben kennen Sie das bestimmt: Die Inventur – wenn im Supermarkt der Warenbestand erfasst wird. Moderne Warenwirtschaftssysteme übernehmen diese Aufgabe heutzutage automatisch – und genau dahin soll auch die Reise der Verwaltungsdaten gehen. Unter einer Dateninventur versteht man den **Prozess der Recherche** nach vorhandenen (Roh-)Daten sowie deren

**Erfassung und Beschreibung.** Sie erfassen ausschließlich Metadaten, also beschreibende Informationen über die vorliegenden Daten. Metadaten enthalten neben einem bezeichnenden Titel typischerweise eine kurze Beschreibung

- was der Datensatz beinhaltet
- wer die Daten hält und
- wo die Daten liegen.

Die Dateninventur ist ein Prozess, im Ergebnis erstellen Sie so einen **Datenkatalog** für Ihre Organisation.

## Ziel der Handreichung

Zur Vorbereitung einer Dateninventur skizziert das Open-Data-Team der byte wichtige Aspekte und Schritte, damit Sie gleich loslegen können. Lesen Sie hier:

- warum es wichtig ist, ein Team zu bilden
- welche Hilfsmittel benötigt werden
- wie Daten am besten beschrieben werden
- welche Datenformate im Hinblick auf die Nachnutzung von Daten interessant sind
- wann ein guter Zeitpunkt ist zu starten und
- wie Sie über eine Dateninventur kommunizieren sollten

Ein Glossar am Ende des Dokuments erklärt einige der erwähnten Begriffe.

# Durchführung einer Dateninventur

Der hier beschriebene Ansatz soll Ihnen helfen pragmatisch eine Dateninventur durchzuführen und einen Datenkatalog aufzubauen. Folgende Punkte dienen der **Vorbereitung der Dateninventur**:

## 1. Dateninventur als Teamarbeit

Die Durchführung der Dateninventur und der Aufbau des Datenkatalogs sind Teamarbeit. Auch wenn Sie das Thema zunächst allein angehen, werden Sie im Laufe der Zeit Mitstreitende finden, die gemeinsam mit Ihnen die Idee eines Datenkatalogs vorantreiben. So kann das Wissen über mögliche Datenquellen, die über mehrere Bereiche der Verwaltung verstreut vorliegen, geteilt werden. Wenn Sie im Team starten, empfiehlt es sich eine zentrale Ansprechperson zu benennen. Nach der Erstellung eines Datenkatalogs in der ersten Version empfehlen wir weitere feste Rollen zuzuweisen, um das Thema langfristig in der Verwaltung zu verankern.

## 2. Tools und Hilfsmittel

Starten Sie mit einer einfachen, zugänglichen Lösung. Es reicht eine übersichtliche Tabelle z.B. in Excel, die später jederzeit in ein professionelles Tool überführt werden kann, wenn der Datenkatalog in Ihrer Institution akzeptiert, genutzt und gelebt wird. Ein Datenkatalog kann viele Formen und Professionalitätsstufen einnehmen, bis hin zu Experten-Tools in

welchen Datenbeziehungen gepflegt und Schnittstellen zu den tatsächlichen Datensätzen etabliert werden können. Um schnell zu starten, finden Sie in unserem Portal eine beispielhafte Excel-Vorlage:

[open.bydata.de/info/inventory](https://open.bydata.de/info/inventory)

## 3. Metadaten für eine strukturierte Erfassung

Definieren Sie vor dem Start der Dateninventur, welche Informationen über die Daten erhoben werden sollen und generieren Sie ein verbindliches Metadaten-Set, also ein **Minimum an Beschreibungen und Attributen** (Metadaten), die erfasst werden sollen. Dies ist besonders wichtig für eine strukturierte Erfassung und das spätere Auffinden der Daten. Wir empfehlen Ihnen, dass Sie die Definition Ihres Metadaten-Sets am deutschen Metadaten-Standard [DCAT-AP.de](https://www.dcat-ap.de/) orientieren<sup>1</sup>. So können entsprechende Katalogeinträge später gut in zentrale Systeme eingespielt werden.

## Rohdaten und Metadaten

Unter Daten bzw. Datensätzen werden hier **Rohdaten** verstanden. Es geht nicht um fertige Datenanalysen wie ausformulierte statistische Jahresberichte sondern um unbearbeitete, ungefilterte Daten. Zur Beschreibung dieser Rohdaten dienen **Metadaten**, also formale und inhaltliche Informationen über Daten.

Folgende Informationen sollten Sie mindestens aufnehmen:

- **Titel:** Bezeichnung des Datensatzes
- **Beschreibung:** Informationen zum Inhalt des Datensatzes, zur Datenstruktur, wichtige Attribute wie Zeitpunkt der Datenerhebung
- **Datenliefernde Stelle:** Verzeichnen der für die Daten zuständigen Organisationseinheit
- **Kategorien:** thematische Einordnung mit Hilfe fester Kategorien – z.B. aus dem [Musterdatenkatalog](#) der Bertelsmann-Stiftung oder nach der [genormten Kategorienliste der EU](#).<sup>2</sup>
- **Datum:** wann wurde der Datensatz zuletzt aktualisiert
- **Kontaktdaten:** Erfassen der datenverantwortlichen Person mind. mit einem E-Mail-Kontakt - für interne Kataloge können diese personalisiert sein; besser eignen sich Funktionspostfächer
- **Verschlagwortung:** aussagekräftige Begriffe, die die Suche nach Daten vereinfachen
- **Lizenz:** Klärung der Nachnutzungsrechte – für eine potenzielle spätere Veröffentlichung der Daten unabdingbar.

#### 4. Datenformate im Fokus

Ziel der Dateninventur ist die Erfassung von Datensätzen und deren mögliche Weiterverarbeitung oder Nachnutzung. Daher ist ein Überblick über vorhandene **Datenformate** von großer Bedeutung. Vor allem **maschinenlesbare Formate** oder auch verfügbare Schnittstellen, die den Zugriff auf Datenbanken ermöglichen, bspw. Web Map Services (WMS), sollten unbedingt ermittelt und für den Datenkatalog erfasst werden. Hier einige Beispiele für maschinenlesbare Formate:

	Formatbeispiele
<b>Strukturierte Textdateien</b>	CSV, JSON
<b>Geodaten</b>	GeoJSON
<b>Karten</b>	WMS, WMF

PDFs eignen sich nur, wenn sie in Kombination mit einem maschinenlesbaren Datensatz vorliegen. Denn: Ein Datenkatalog ist kein Schriftenverzeichnis.

Wenn Sie bei der Dateninventur auf **Datenserien** oder sehr fein untergliederte Datensätze stoßen, also z.B. Daten, die regelmäßig in gleicher Struktur erhoben, erfasst und evtl. schon veröffentlicht werden, prüfen Sie hier, ob es sinnvolle Zusammenfassungen der Datensätze geben kann. Können bspw. statistische Daten zu Einwohnerzahlen in Jahresdatensätzen zusammengefasst werden, statt einzelner Monatszahlen nachzuweisen?

## 5. Projektstart Dateninventur

Es ist hilfreich, die Dateninventur mit einem Projekt bspw. der Erarbeitung einer Digitalisierungsstrategie zu koppeln oder daraus ein eigenes internes Projekt zu initiieren. Auch laufende thematisch eingegrenzte Projekte eignen sich für eine erste Dateninventur. Wenn Sie bei Ihnen z.B. eine neue datenbasierte Anwendung ansteht – wie der Aufbau eines Klimadashboards – kann es sinnvoll sein, in diesem Rahmen oder Vorfeld eine Dateninventur durchzuführen, um vorhandenen Datenquellen wie Wetterdaten, Sensordaten etc. zu ermitteln.

Folgende Aspekte können gute Anknüpfungspunkte für eine Dateninventur sein:

- **Wiederkehrende Fragestellungen:** Ihre Pressestelle kennt die wiederkehrenden Fragestellungen der Medien oder der Bevölkerung, die sich mit einem gepflegten Datenkatalog schneller beantworten lassen würden.
- **Erfassung durch die Fachabteilungen:** Das Know-how über die im Rahmen der Leistungserbringung benötigten bzw. erstellten Daten ist in den Fachabteilungen vorhanden. Sofern ein Produkt- und Leistungskatalog der Verwaltung vorliegt, kann dieser als Ausgangspunkt herangezogen werden.
- **Entlang der IT-Systeme:** Blicken Sie systematisch auf IT-Systeme (Datenbanken, Applikationen, Fachanwendungen, Register), in denen Daten zur Bearbeitung erfasst werden.
- **Musterdatenkataloge als Orientierung:** Blick in Musterdatenkataloge, bspw. [Musterdatenkatalog der Bertelsmann Stiftung](#). Hier finden Sie Hinweise, welche weiteren Daten in Ihrer Organisation vorliegen könnten.
- **Formulare durchsehen:** Zusätzlich kann es hilfreich sein, in einem Verfahren benötigte Antragsformulare durchzusehen. Die dort abgefragten Daten geben oft sehr konkrete Hinweise auf vorhandene Daten.

## 6. Kommunikation ist alles.

Die Motivation, sich einer Dateninventur zu widmen und einen Datenkatalog zu erarbeiten, kann unterschiedlicher Art sein. Ein Anwendungsfall ist die Vorbereitung von Verwaltungsdaten zur Veröffentlichung als Open Data. Andere Treiber können Digitalisierungsprojekte sein, welche bspw. Transparenz stärken sollen (z.B. Fortschritts-Tracking bei Bauanträgen) oder die Automatisierung vorantreiben (z.B. automatisierte Analysen von Verkehrsströmen oder Bürgerinformation auf Basis geänderter Einwohnermeldedaten). Dennoch: Auf das Vorhaben eine Dateninventur durchzuführen, wird häufig mit Rückfragen und Bedenken reagiert. Bereiten Sie ausgehend von Ihrer Motivation Argumente vor, die das Vorgehen begründen und finden Sie so Mitstreiter.

- **Allgemeingültige Motivation:** Durch einen fertigen Datenkatalog können Nachfragen durch Bürgerinnen und

Bürger oder Journalisten künftig mit weniger Aufwand beantwortet werden.

- **Mehrwerte erzählen:** Überlegen Sie, welche Personen oder Abteilungen stark von einem Datenkatalog profitieren würden und erzählen Sie diese Geschichte in einer Abteilungsbesprechung.
- **Bedenken aufgreifen:** Wenn Ihnen Bedenken entgegengetragen werden – nehmen Sie diese ernst. Datenschutz, kritische Infrastrukturen und IT-Sicherheit lassen Bedenken laut werden. Hilfreich ist die Unterstützung des Datenschutzbeauftragten bzw. des IT-Security-Verantwortlichen einzuholen.
- **Offizieller Auftrag:** Die Dateninventur ist im besten Fall ein offizielles Projekt, das bspw. Durch die Verwaltungsleitung beauftragt wird.

Generell gilt: Je mehr Menschen Bescheid wissen und ein positives Bild einer Dateninventur haben, desto einfacher wird es, weitere Datenbestände zu identifizieren und für das Thema zu begeistern. Daher sollten regelmäßig Informationen geteilt werden – das können Info-Mails, Intranet-Einträge zum Fortschritt sein oder direkte Einladungen zum Mitmachen.

## Aktualität des Datenkatalogs

Ein Datenkatalog ist im Grunde nie fertig, sondern ein **Nachschlagewerk**, das von seiner Nutzung ebenso profitiert wie von regelmäßigen Aktualisierungen und Anpassungen. Achten Sie also darauf, den Datenkatalog aktuell zu halten und wenn nötig anzupassen. Dafür sollten mittelfristig klare Prozesse und Rollen bestimmt werden, die dafür sorgen, dass der Datenkatalog seine Aktualität behält. Eine pragmatische erste Möglichkeit dafür ist es, quartalsweise Update-Mails einzurichten, mit der Aufforderung, Metadaten auf ihre Richtigkeit zu prüfen, nicht mehr existente Datenbestände abzumelden und neue Datenbestände anzumelden.

## Zugriffsrechte und Veröffentlichung als Open Data

Nach erfolgreichem Aufbau eines Datenkatalogs stellt sich die Frage, wer auf diesen zugreifen können soll. Hier ist zu unterscheiden zwischen rein verwaltungsinternen Datenbeständen und Daten, die generell oder in anonymisierter Form geteilt oder veröffentlicht werden können. An dieser Stelle kommt das **Potenzial von Open Data** wieder ins Spiel. Die byte hat mit dem Open-Data-Portal Bayern eine zentrale Infrastruktur geschaffen, in der Sie Ihre Datensätze einfach und schnell veröffentlichen können. Denn:

Sie haben nun einen guten Überblick, welche Daten in Ihrer Organisation wo vorhanden sind und von wem gepflegt werden, es werden sich bereits bei der Erfassung intern erste

**Synergie-Effekte** über Referats- und Abteilungsgrenzen hinweg ergeben haben – nutzen Sie dieses Netzwerk und gehen Sie den nächsten Schritt: Veröffentlichen Sie die Datensätze, die frei zugänglich sein sollten, um neue Angebote für Wirtschaft, Forschung und Zivilgesellschaft zu schaffen.

Offene Daten ermöglichen beispielsweise Unternehmen, ihre Geschäftsmodelle zu verbessern und stärken so die Wirtschaft vor Ort. Auch die Zivilgesellschaft wird gestärkt, indem Sie Verwaltung transparent machen und aktive Zusammenarbeit fördern. In einer datengetriebenen Welt machen Sie mit Datenveröffentlichungen auch sich selbst sichtbar. Je mehr maschinenlesbare, hochqualitative Daten aus Ihrer Verwaltung verfügbar sind, desto einfacher können diese Daten z.B. in Künstlichen Intelligenzen wie ChatGPT verarbeitet werden. Schließen Sie sich anderen Behörden, Ministerien und Kommunen an und veröffentlichen auch Sie Ihre Daten gerne auf dem bayerischen Portal [open bydata](#) oder als Kommune mit einer lokalen Open-Data-Präsenz. Über eine automatische Schnittstelle werden Ihre offenen Daten so in [GovData](#), das deutsche Open-Data-Portal, und [data.europa.eu](#), das Portal der EU, eingespielt. Werden Sie mit der byte als Partner Teil der Open-Data-Community.

## Wir für Sie

Wenn Sie mehr zum Thema Open Data in Bayern und die Möglichkeiten der Veröffentlichung Ihrer Datensätze im Open-Data-Portal [open bydata](#) erfahren möchten, nehmen Sie gerne mit uns Kontakt auf. Als Open-Data-Team der byte unterstützen wir Sie in der bayerischen Verwaltung dabei, Daten zu teilen und Mehrwerte zu schaffen. Denn: Offene Daten schaffen die Basis für innovative Produkte und Services.

Melden Sie sich bei uns, wir betrachten gerne Ihren individuellen Fall und entwickeln mit Ihnen eine Lösung.

Wir freuen uns auf Ihre Nachricht unter [openbydata@byte.bayern](mailto:openbydata@byte.bayern)

## Weiterführende Links

- [Informationsseiten auf open bydata](#)
- [Musterdatenkatalog der Bertelsmann-Stiftung](#)
- [Datenblätter ODIS Berlin](#)

## Nachnutzung

Wir freuen uns, wenn unsere Handreichung nachgenutzt wird! Nutzen Sie die Texte und Informationen daraus gerne weiter.



[CC BY 4.0](#) „open bydata, byte - Bayerische Agentur für Digitales GmbH“

<sup>1</sup> DCAT-AT.de ist der formale Austauschstandard für offene Verwaltungsdaten. Auch das bayerische Open-Data-Portal open bydata nutzt DCAT-AP.de. Wenn Sie Datensätze zukünftig über das Portal veröffentlichen wollen, empfiehlt es sich, diese gleich unter Nutzung des deutschen Metadatenstandards zu erfassen.

<sup>2</sup> Im Hinblick auf eine spätere Veröffentlichung empfiehlt sich die EU-Kategorie, die sich als Standard für Open Data etabliert haben. Thematisch weiteraufgestellt und somit gut für einen Einstieg, welche Daten vorhanden sein können, ist der Musterkatalog der Bertelsmann-Stiftung.

## Glossar

Begriff	Erläuterung
API-Schnittstelle	Eine API (Application Programming Interface) ist eine Schnittstelle, die es unabhängigen Anwendern ermöglicht, miteinander zu kommunizieren und Daten auszutauschen.
Datenbestand	Als Datenbestand wird die Gesamtheit der vorhandenen Daten in Bezug auf einen Kontext bezeichnet. Ein kommunaler Kontext könnten Einwohnermeldedaten sein. Der dazugehörige Datenbestand wären alle Daten zu diesem Thema.
Datenformat	Datenformat bezieht sich auf die Struktur, in der Daten gespeichert werden. Es gibt viele verschiedene Datenformate, wie z.B. Textdateien (.txt) oder Bilddateien (.jpg, .png).
Datensatz	Ein Datensatz ist eine Gruppe von inhaltlich zusammenhängenden und zu einem Objekt gehörenden Datenfeldern, zum Beispiel ein Excel File.
<a href="https://www.dcat-ap.de/">DCAT-AP.de</a>	Das deutsche Metadatenmodell zum Austausch von offenen Verwaltungsdaten – auch bez. als Metadatenstandard – dient der einheitlichen strukturierten Erfassung von Daten.
GIS	Ein geografisches Informationssystem (GIS) bezeichnet ein System zur Erfassung, Verwaltung und Analyse von Geo-Daten.
Lizenzen	Über Lizenzen werden die Nutzungsrechte an Daten geklärt, sie regeln in welcher Art Daten rechtssicher weiterverarbeitet werden dürfen. Im Kontext von Open Data sind vor allem sogen. Creative Commons Licenses und Datenlizenzen von Relevanz.
Maschinenlesbarkeit	Maschinenlesbare Daten sind von Maschinen/Computern les- und interpretierbar. Eine typische Prüffrage wäre, ob man die vorliegenden Daten von Software durchsuchen lassen kann, was mit einer Excel-Datei sehr gut geht, bei Bildern aber, wenn überhaupt, nur sehr eingeschränkt.

Metadaten	Metadaten sind Daten über Daten, zum Beispiel das Bearbeitungsdatum einer Datei oder eine thematische Kategorie wie "Umwelt".
Open Data	Als Open Data werden Daten bezeichnet, die mit einer offenen Lizenz im Internet bereitgestellt werden und von allen zu jedem Zweck genutzt, weiterverbreitet und weiterverwendet werden dürfen.
Rohdaten	Rohdaten sind ungeprüfte und unbearbeitete Daten, wie sie von einer Datenquelle geliefert werden. Sie liegen in dem Format vor, wie es von der jeweiligen Datenquelle bereitgestellt wird.
Strukturierte Daten	Strukturierte Daten sind Daten, die in einem vorgegebenen Format strukturiert wurden, bevor man sie im Datenspeicher abgelegt hat. Ein Beispiel hierfür sind Excel-Tabellen.
Unstrukturierte Daten	Unstrukturierte Daten sind Daten, die in einem nativen Format gespeichert und erst dann bearbeitet werden, wenn sie verwendet werden. Beispiele hierfür wären Social Media Posts, Bilder oder E-Mails.