



# Open Data Literacy

Ein Leitfaden für Behörden und Unternehmen

### Herausgeber

Bitkom e. V.  
Albrechtstraße 10  
10117 Berlin  
T 030 27576-0  
bitkom@bitkom.org  
www.bitkom.org

### Ansprechpartner

Dr. Frank Termer | Bitkom e.V.  
T 030 27576-232 | f.termer@bitkom.org

### Verantwortliches Bitkom-Gremium

AK Open Data / Open API

### Projektleiter

Michael Ochs | Fraunhofer Institut für Experimentelles Software Engineering IESE

### Layout

Anna Stolz | Bitkom e.V.

### Titelbild

© Jason Pofahl – unsplash.com

### Copyright

Bitkom 2023

Diese Publikation stellt eine allgemeine unverbindliche Information dar. Die Inhalte spiegeln die Auffassung im Bitkom zum Zeitpunkt der Veröffentlichung wider. Obwohl die Informationen mit größtmöglicher Sorgfalt erstellt wurden, besteht kein Anspruch auf sachliche Richtigkeit, Vollständigkeit und/oder Aktualität, insbesondere kann diese Publikation nicht den besonderen Umständen des Einzelfalles Rechnung tragen. Eine Verwendung liegt daher in der eigenen Verantwortung des Lesers. Jegliche Haftung wird ausgeschlossen. Alle Rechte, auch der auszugsweisen Vervielfältigung, liegen beim Bitkom.

	Danksagung	5
	Vorwort	6
1	<b>Motivation für einen Leitfaden Data Literacy für Open Data</b>	7
2	<b>Data Literacy, Daten und Wissen</b>	10
	Data Literacy kurz erklärt	10
	Einordnung von Daten im Kontext Wissenserzeugung	12
3	<b>Open Data Fähigkeiten für Behörden und Unternehmen</b>	15
	Open Data Literacy: Die Phasen im Lebenszyklus	15
	Data Literacy Fähigkeiten von Data Provider und Consumer Organisationen entlang der Phasen im Open Data Lifecycle	18
	Data Literacy: Rollen in Open Data-Projekten und deren Kenntnisse	23
4	<b>Szenarien für Open Data mit professioneller Data Literacy</b>	26
	Szenario 1: Parkplatzdaten auf Autobahnparkplätzen	26
	Szenario 2: Abwasser-Monitoring als Frühwarnsystem bei drohenden Corona-Ausbrüchen	27
5	<b>Zusammenfassung</b>	30

1	Der Strandticker für die Lübecker Bucht	8
2	Daten im Kontext Wissenserzeugung	12
3	Beispiel aus dem Kontext Hausautomatisierung (Heizungssteuerung)	13
4	Sechs Phasen des Open Data Lebenszyklus und Fähigkeiten je Phase	17
5	Welche Rolle ist in welcher Phase aktiv?	25
6	Dashboard zum Abwasser-Monitoring von SARS-CoV-2 im Landkreis Berchtesgadener Land [Stand: Juni 2023]	29
	<a href="#">Tabelle 1: Fähigkeiten im Open Data Lebenszyklus</a>	19
	<a href="#">Tabelle 2: Open Data-bezogene Rollen in Provider und Consumer Organisationen</a>	24

# Danksagung

■ Erfolg entsteht nicht allein durch die individuelle Anstrengung, sondern durch die Zusammenarbeit und das Engagement eines Teams.

Ein herzliches Dankeschön geht an alle, die an der Erstellung des Leitfadens »Open Data Literacy« mitgewirkt haben. Nach 15 Monaten intensiver Arbeit ist der Leitfaden endlich fertiggestellt. Trotz Herausforderungen und Verzögerungen hat die Gruppe ihre Ziele erreicht.

Die Autorinnen und Autoren haben einen unschätzbaren Beitrag geleistet. Ihr Fachwissen, Engagement und Entschlossenheit haben dazu beigetragen, dass der Leitfaden unterhaltsam und informativ geworden ist. Die unzähligen Webkonferenzen und Abstimmungen waren eine Herausforderung, aber dank der Zusammenarbeit konnte das Projekt erfolgreich abgeschlossen werden.

Unser Dank gilt besonders dem Vorstand des Arbeitskreises »Open Data / Open API«, der das Projekt von Anfang an unterstützt hat. Durch ihr Vertrauen und ihre Visionen konnte der Leitfaden für Organisationen aus Wirtschaft und Verwaltung veröffentlicht werden, um sie bei der Bereitstellung und Nutzung von Daten als Open Data zu unterstützen.

Besonders hervorzuheben ist die Arbeit des Projektleiters Michael Ochs. Sein Engagement, Vorbild und Fachwissen haben zum erfolgreichen Abschluss des Projekts beigetragen. Durch seine Koordination und Beharrlichkeit hat er das Team motiviert und geführt.

Data Literacy für Open Data ist von großer Bedeutung. In einer datenorientierten Gesellschaft ist es entscheidend, dass Organisationen und Einzelpersonen über die notwendigen Fähigkeiten verfügen, um Daten zu verstehen und effektiv zu nutzen. Der Leitfaden »Open Data Literacy« trägt dazu bei, diese Kompetenz zu fördern und den Fortschritt voranzutreiben.

Wir danken den Autorinnen und Autoren für das notwendige Wissen, das in einem kompakten Leitfaden zusammengefasst wurde.

- Michael Binzen, DB System GmbH
- Kapitalina Hänsel, Toll Collect GmbH
- Marie Jansen, Capgemini Deutschland GmbH
- Michael Ochs, Fraunhofer Institut für Experimentelles Software Engineering – IESE
- Martin Schuh, Bundesanzeiger Verlag GmbH
- Kerstin van de Sand, Esri Deutschland GmbH

# Vorwort

Der Leitfaden »Open Data Literacy« richtet sich an Organisationen aus Wirtschaft und Verwaltung, die Daten als offene Daten bereitstellen oder nutzen möchten. Die Verfasserinnen und Verfasser dieses Leitfadens sehen Data Literacy als integralen Teil einer Datenkultur, die in der Gesellschaft in vielen Belangen das Erzielen von Mehrwerten ermöglicht. Dabei ist es den Verfasserinnen und Verfassern ein Anliegen, dass die Bereitstellung von Open Data nicht nur die reine Erfüllung regulatorischer Anforderungen sein darf, sondern der Nutzen für die Allgemeinheit im Vordergrund steht. Mit offenen Daten kann ein datenorientierter Mehrwert für das Gemeinwohl geschaffen werden, aber gleichzeitig auch im direkten Umfeld des Datenbereitstellers ein unmittelbarer Effekt erzielt werden. Damit wird Mehrwert und Innovation für alle spür- und nutzbar. Allerdings muss dazu kompetent mit offenen Daten umgegangen werden – etwas, das noch nicht bei allen Zielgruppen selbstverständlich ist. Die Kompetenz, mit Daten und insbesondere offenen Daten umzugehen, wird zur Voraussetzung, um in Zukunft in einer datenorientierten Gesellschaft das volle Potenzial von Daten zu nutzen. Data Literacy ist somit die wichtige Grundlage, um das Wertschöpfungspotenzial in Daten einerseits zu verstehen und andererseits auch konsequent nutzen zu können.

Der Leitfaden richtet sich dabei an Entscheiderinnen und Entscheider und operative Mitarbeitende in Behörden und Unternehmen, sodass insgesamt alle Stakeholder von Open Data – Open Data Erzeuger, Open Data Bereitsteller und Open Data Konsumierende bzw. Nutzerinnen und Nutzer – gleichermaßen adressiert werden.

Der Leitfaden möchte daher zunächst deutlich machen, was Data Literacy (dt. Datenkompetenz) ist und welche Fähigkeiten essenziell sind, um den Umgang mit Daten zu stärken, gelungene Open Data Angebote zu erzeugen und diese optimal und nutzbringend zu verwenden. Die Fähigkeiten, die Data Literacy ausmachen, werden im Umfeld von Open Data und Open API abgebildet und praxisnah erklärt. Der Leitfaden wird durch zwei Szenarien von Open Data abgerundet, die neue Perspektiven auf Open Data und deren Quellen bieten, sodass der gesamtgesellschaftliche Vorteil von Open Data deutlich wird.

Dieser Leitfaden behandelt Data Literacy auf der organisatorischen Ebene. Dabei liegt der Fokus auf fachlichen Themen wie z. B. einem Ablauf von der Identifikation über die Bereitstellung bis hin zur Nutzung von Open Data in digitalen Services sowie den dazu erforderlichen Fähigkeiten der Organisationen und Personen, die sich mit Open Data befassen. Auf der tiefer liegenden technischen Ebene spielen weitere Faktoren eine wichtige Rolle, um Open Data Projekte erfolgreich durchführen und verstetigen zu können. Dazu gehört aus dem Blickwinkel der Digitalisierung vor allem das Thema Maschinenlesbarkeit von Daten.

# 1. Motivation für einen Leitfaden Data Literacy für Open Data

Data Literacy (dt. Datenkompetenz) bezeichnet die Fähigkeiten von Organisationen und Individuen, mit Daten entlang ihres Lebenszyklus von der Erfassung bis hin zur Nutzung von Daten kompetent umzugehen<sup>1</sup>. Datenkompetenz ist somit auch eine wesentliche Voraussetzung für die erfolgreiche Digitalisierung von Daten und setzt ein gewisses Level an Digitalkompetenzen voraus.

Eine Vielzahl an Behörden und Unternehmen ist bislang noch nicht im digitalen (Daten-)Zeitalter angekommen. Dies macht sich durch zahlreiche Defizite und Herausforderungen bemerkbar. Ohne die entsprechenden Kompetenzen wird zum Teil verhindert, dass Open Data überhaupt oder in einer für Dritte nutzbringenden Form bereitgestellt werden. Die Bedenken und Hürden, Daten offen und flächendeckend bereit zu stellen sind dabei vielfältig<sup>2</sup>.

Zu den häufigsten Bedenken zählen z. B. der Schutz sensibler Daten, Vermeidung von Wettbewerbsvorteilen für Konkurrenten, Verlust der Kontrolle und Deutungshoheit über die Daten oder eine unklare bzw. fehlende Perspektive hinsichtlich der Nutzeffekte der betreffenden Daten.

Dabei liegen die Vorteile und Potenziale von Open Data klar auf der Hand: Sie können großen gesellschaftlichen und wirtschaftlichen Nutzen stiften und zu vielen für die Gestaltung der Gegenwart und der Zukunft relevanten Zielen beitragen. Hier sind z. B. Nachhaltigkeitsthemen zu nennen. Eine Verkehrswende, bessere Klimaverträglichkeit sowie Erleichterungen im Alltag durch mobile Applikationen und Kartenanwendungen, wie z. B. Echtzeitfahrpläne öffentlicher Verkehrsmittel, Verkehrsdichte im Straßennetz, Auslastung von Parkplätzen oder die Auslastung von Strandbereichen an der Ostsee in der Urlaubssaison könnten durch Open Data aus adäquaten und verlässlichen Quellen bei zielorientierter Nutzung ermöglicht werden und so zu gesellschaftlichem Mehrwert beitragen. Ein Beispiel für einen digitalen Service, basierend auf Open Data ist der »Strandticker«, der Strandgänger bei der Suche nach dem für sie besten Strandplatz durch eine App mit Informationen über die Auslastung der Strände und der am Strand gelegenen Parkplätzen in der Lübecker Bucht unterstützt. Die Strandbelegung wird dabei mittels einer Ampel und die freien Parkplätze durch deren absolute Anzahl visualisiert. Abbildung 1 zeigt einen beispielhaften Screenshot des Strandtickers.

1 ↗ <https://www.bigdata-insider.de/was-ist-data-literacy-a-823501/>

2 ↗ <https://www.bitkom.org/sites/main/files/file/import/171103-Open-Data.pdf>

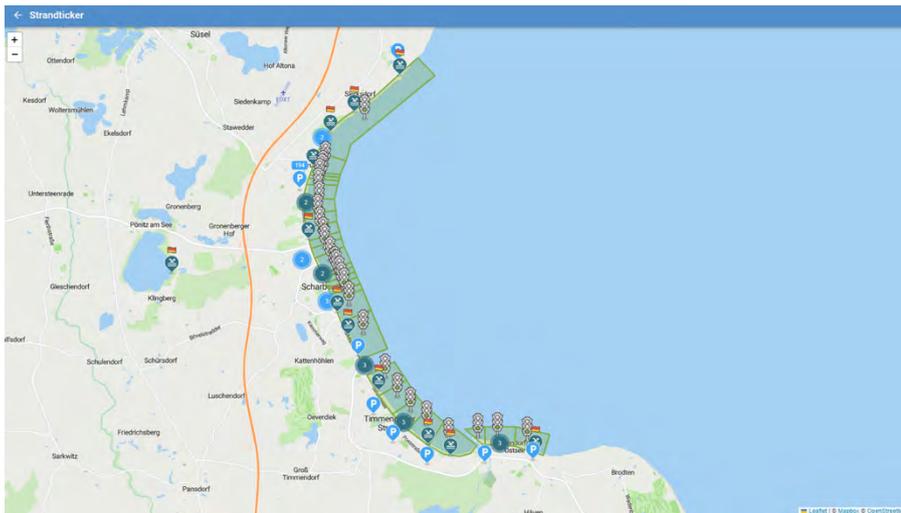


Abbildung 1: Der Strandticker für die Lünecker Bucht<sup>3</sup>

Hierfür müssen Hürden überwunden werden und es muss erkannt werden, dass Dritte Daten nur veredeln können und zusätzlichen Nutzen stiften, wenn Daten (unter Beachtung geltender Gesetze) offengelegt werden. Dies bedeutet, dass eine Offenheit für Open Data in den Köpfen von Entscheidungsträgerinnen und -trägern etabliert werden muss: Data Literacy muss in Behörden und Unternehmen eingeführt und dort verankert werden. Data Literacy bezeichnet also die Fähigkeit, mit Daten umgehen zu können und Daten zu öffnen. Die gesetzliche Verpflichtung für Behörden und öffentliche Verwaltung, Open Data bereitzustellen, kann mithilfe von Data Literacy-Fähigkeiten besser und schneller realisiert werden. Data Literacy stellt damit eine Weichenstellerin dar, um die im Gesetz geforderte Offenheit für Verwaltungsdaten erfolgreich praktizieren zu können. Sie ist notwendige Grundlage des Handelns. Darüber hinaus müssen Daten als Grundlage der Digitalisierung verstanden werden.

Im Koalitionsvertrag 2021<sup>4</sup> sind die Einführung eines Rechtsanspruches auf Open Data und der Wille, die Datenexpertise öffentlicher Stellen zu verbessern verankert. In Bezug auf Forschungsdaten wird über die Datenteilung von vollständig anonymisierten und nicht personenbezogenen Daten für Forschung im öffentlichen Interesse gesprochen. Diese zu ermöglichen, wird als Ziel gesetzt (siehe Seite 14, Nutzung von Daten und Datenrecht; Seite 18, Forschungsdaten). Zur Erreichung dieser Ziele wollen wir in Form dieses Leitfadens einen Beitrag leisten und im Folgenden einige für Open Data relevante Bereiche beispielhaft beleuchten und deren potenziellen Nutzen darstellen<sup>5</sup>:

- **Transparenz und demokratische Kontrolle**, z. B. von behördlichen Prozessen oder Regierungstätigkeiten, wie z. B. Daten von Wahlen auf allen demokratischen Ebenen über mehrere Jahre in einem einfach erreichbaren und maschinenlesbaren Datenangebot

3 Quelle: ↗ <https://www.luebecker-bucht.guide/beachticker/map>, abgerufen am 28. Juni 2023, 12:39 Uhr

4 ↗ [https://www.spd.de/fileadmin/Dokumente/Koalitionsvertrag/Koalitionsvertrag\\_2021-2025.pdf](https://www.spd.de/fileadmin/Dokumente/Koalitionsvertrag/Koalitionsvertrag_2021-2025.pdf)

5 ↗ <http://opendatahandbook.org/guide/de/why-open-data/>

- Partizipation, z. B. Bereitstellung von Umweltmessdaten durch viele einzelne Bürgerinnen und Bürger wie z. B. flächendeckende Luftmessungen, Lärmmessungen
- Befähigung der Bürgerinnen und Bürger zur Schaffung von zusätzlichen Informationsangeboten z. B. durch Crowdsourcing, z. B. in Form von individuellen Datenspenden wie sie bei OpenStreetMap möglich sind
- Verbesserte oder neue Produkte und Dienstleistungen in der Privatwirtschaft, z. B. smarte Services im gesellschafts- und umweltpolitischen Umfeld auf Basis von Verkehrs- oder Wetterdaten
- Innovationen, die durch Bereitstellung neuer Quellen für offene Daten zu völlig neuen und nutzenstiftenden Lösungen führen können
- Verbesserte Effizienz und Wirksamkeit von Behördendienstleistungen, z. B. eine Verringerung des Einflusses des Fachkräftemangels in Behörden durch die Intensivierung von Kollaborationen unter Behörden sowie zwischen Behörden und privaten Organisationen und eine Verbesserung der Services gegenüber Bürgerinnen und Bürgern
- Verbesserungen von behördeninternen Abläufen (Verfahren und Prozesse), die sich z. B. aus der Anwendung von Data Driven Government und datenbasierten, informierten Entscheidungen ergeben
- Wirkungsmessung von politischen Entscheidungen, z. B. Wirkung des 9-Euro-Tickets hinsichtlich Entlastung der automobilen Verkehrswege bzw. Staudaten
- Neues Wissen aus einer Kombination von Datenquellen und Mustern aus großen Datensammlungen, um z. B. politische oder unternehmerisch wichtige Entscheidungen optimal unterstützen zu können

Diese Liste von Nutzungspotenzialen könnte noch nahezu beliebig erweitert und ergänzt werden.

Die Sensibilisierung für mehr Offenheit möchte dieser Leitfaden erreichen und aufzeigen, welche Fähigkeiten erforderlich sind, um den Schritt zu Open Data professionell, ohne Bedenken und Ressentiments zielgerichtet und sicher gehen zu können. Aktuelle Studien zeigen, dass etwa jedes zweite Unternehmen (46 Prozent) bei sich einen Mangel in der Steuerung seines digitalen Portfolios sieht und fast zwei von drei Unternehmen (62 Prozent) einen Mangel an technischen Skills in der IT. Beides sind auch zentrale Voraussetzungen für die Bereitstellung von Open Data in für andere nutzbarer Form, genauso wie die Nutzung solcher Daten, um Services anzubieten oder informierte, datenbasierte Entscheidungen zu treffen<sup>6</sup>.

6 [https://content.rolandberger.com/hubfs/07\\_presse/Roland\\_Berger\\_Focus\\_Digital\\_Dilemma\\_N3XT\\_2022.pdf](https://content.rolandberger.com/hubfs/07_presse/Roland_Berger_Focus_Digital_Dilemma_N3XT_2022.pdf)

## 2. Data Literacy, Daten und Wissen

### 2.1. Data Literacy kurz erklärt

Der Begriff Data Literacy (Datenkompetenz) wurde in der Motivation bereits grob umrissen. Für Data Literacy existieren eine Reihe von Definitionen mit zum Teil sehr unterschiedlichen Detaillierungsgraden.

Data Literacy ist die Fähigkeit von Organisationen und Individuen, Daten im Kontext zu lesen, zu schreiben und zu kommunizieren, mit einem Verständnis der Datenquellen und Konstrukte, der angewandten Analysemethoden und -techniken und der Fähigkeit, den Anwendungsfall, den daraus resultierenden Geschäftswert und den erzielten Effekt zu erkennen und zu beschreiben<sup>7</sup>.

Sie umfasst dabei Fähigkeiten, Daten auf kritische Art und Weise zu sammeln, zu managen, zu bewerten und anzuwenden. Wenn Daten Entscheidungsprozesse, so wie in der Motivation bereits exemplarisch benannt, unterstützen sollen, braucht es kompetente Antworten auf mehrere grundlegende Fragen<sup>8</sup>:

- **Welche Daten in meiner Organisation könn(t)en als Open Data Nutzen stiften?**  
Die Existenz von Daten in der eigenen Organisation und das Kennen der Bedarfe für datenbasierte Dienste und Informationen innerhalb und außerhalb der eigenen Organisation – auch jenseits regulatorischer Anforderungen – bilden den Grundstein für Überlegungen und Aktionen zu Open Data-Angeboten.
- **Was will ich mit Daten machen?** Daten und Datenanalysen sind kein Selbstzweck, sondern dienen einer konkreten Anwendung in der realen Welt.
- **Was kann ich mit Daten machen?** Datenquellen und deren Qualität sowie der Stand der technischen und methodischen Entwicklungen eröffnen Möglichkeiten und setzen Grenzen.
- **Was darf ich mit Daten machen?** Alle gesetzlichen Regeln der Datennutzung (z. B. Datenschutz, Urheberrechte und Lizenzfragen) müssen immer mitbedacht werden.
- **Was soll ich mit Daten machen?** Weil Daten eine wertvolle Ressource darstellen, leitet sich daraus ein normativer Anspruch ab, sie zum Wohl von Individuen und Gesellschaft zu nutzen.

7 ↗ <https://www.gartner.com/en/information-technology/glossary/data-literacy>

8 ↗ <https://www.stifterverband.org/charta-data-literacy> und Ergänzungen der Arbeitsgruppe Data Literacy

- Was kann passieren, wenn ich Daten mit hohem Open Data-Potenzial nicht für Dritte nutzbar bereitstelle? Open Data eröffnen beachtlichen volkswirtschaftlichen Mehrwert und bilden die Basis für die Digitalisierung von Verwaltung, Wirtschaft, Wissenschaft und Zivilgesellschaft. Daten mit hohem Open-Data-Potential nicht »open« zu stellen, hieße diese gesamtgesellschaftlichen Mehrwerte nicht zu nutzen.

Technisch bedeutet Data Literacy eine Reihe von Fähigkeiten, die Organisationen und Individuen kennen und einsetzen müssen, um sowohl regulatorische als auch Mehrwertziele zu erreichen. Dies sind auf der obersten Ebene das Erkennen von Datenbeständen und das Antizipieren von Bedarfen, die diesen Daten gegenüberstehen können. Die Fähigkeit, Daten zu erstellen, setzt eine etablierte Datenkultur voraus. Hier ist neben der reinen regulatorischen Anforderung Open Data bereitzustellen auch ein Verständnis für den Nutzen der Daten über die Grenzen der eigenen Organisation hinaus essenziell. Nur so kann eine Erstellung der Daten aus relevanten Datenquellen nutzenorientiert erfolgen. Dabei kann die Häufigkeit der Erstellung von einmalig über eine regelmäßige Erstellung bis hin zu einer Echtzeiterstellung (bspw. über offene Application Programming Interfaces, sog. APIs) der Daten beim Entstehen neuer Werte variieren. An die Erzeugung von Open Data schließt sich die Bereitstellung der Daten selbst an. Die technische Bereitstellung von Daten kann bereits von einer Organisation übernommen werden, die selbst nicht die Erstellung der Daten durchgeführt hat. Die Nutzung der Daten schließt sich an deren Bereitstellung an. Die Fähigkeit der Nutzung von Daten umfasst die Aufbereitung, Auswertung, Analyse und Interpretation der Daten und im Ergebnis die daraus resultierende Ableitung von Entscheidungen und Handlungen. So erst kann ein Effekt aus der Beschäftigung mit Daten erzielt und statt einem reinen Selbstzweck (oder aus regulatorischen Gründen) ein unternehmerischer oder gesellschaftlicher Nutzen erreicht werden<sup>9</sup>.

9 ↗ <https://de.wikipedia.org/wiki/Datenkompetenz>

## 2.2. Einordnung von Daten im Kontext Wissenserzeugung

Der Begriff des Wissens und wie es zustande kommt, findet in der Literatur nicht immer eine offensichtliche Antwort. Bei der Ableitung von Wissen aus dem Hinzuziehen von Daten findet sich ein recht einfaches Modell:

»Typischerweise werden Informationen in Bezug auf Daten, Wissen in Bezug auf Informationen, und Weisheit in Bezug auf Wissen definiert.«<sup>10</sup> Dieser Sachverhalt wird in Abbildung 2 kurz grafisch dargestellt.



Abbildung 2: Daten im Kontext Wissenserzeugung

Sollen Handlungsempfehlungen abgeleitet werden, braucht es verfügbare Daten, aus denen Informationen gewonnen werden können. Aus diesen wiederum kann Wissen generiert werden, welches dabei unterstützt, »Weisheit« abzuleiten. Somit können Handlungsempfehlungen basierend auf einer transparenten Grundlage ausgesprochen werden. Diese Zusammenhänge werden oft in Form einer hierarchischen Darstellung als DIKW-Pyramide<sup>11 12</sup> (»data«, »information«, »knowledge«, »wisdom«) visualisiert. Diese aufeinander aufbauende Form soll die Abhängigkeit der Beziehungen dieser Begriffe tiefer verdeutlichen.

10 ↗ [https://en.wikipedia.org/wiki/DIKW\\_pyramid](https://en.wikipedia.org/wiki/DIKW_pyramid)

11 ↗ [https://en.wikipedia.org/wiki/DIKW\\_pyramid](https://en.wikipedia.org/wiki/DIKW_pyramid)

12 ↗ <http://www.systems-thinking.org/dikw/dikw.htm>

Das folgende Beispiel in Abbildung 3 aus dem Umfeld der Hausautomatisierung soll die beteiligten Kategorien veranschaulichen.

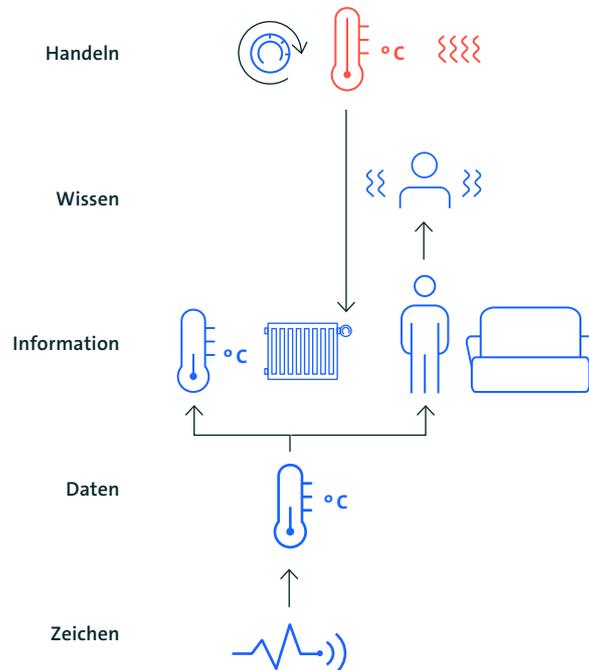


Abbildung 3: Beispiel aus dem Kontext Hausautomatisierung (Heizungssteuerung).

Den Grundstock in diesem Beispiel bildet die Ebene der »Zeichen und Symbole«: Buchstaben, Zahlen und Sonderzeichen. In Kombination mit einem verwendeten Zeichensatz (»encoding«) wird von der Codierungsebene gesprochen.

Im Umfeld der Hausautomatisierung zählen hierzu Sensoren, welche analoge Sachverhalte diskretisieren (und damit digitalisieren).

Darauf aufbauend formen Syntaxregeln Zeichen zu »Daten und Aussagen«. Somit sind diese messbar, sortier- bzw. anordenbar und strukturiert. Werte eines Sensors können durch entsprechende Kalibrierung und gemäß eines Verwendungszwecks z. B. einer Temperatur zugeordnet werden – dieser Sensor zeigt 13 °C an.

Um an »Informationen« zu gelangen, wird Verständnis über die Beziehungen zwischen den Daten benötigt. Verknüpft man diese mittels Semantik, also interpretiert und versteht diese, entstehen Informationen. Der Zuordnung dieses Sensors zu einem Raum kann man eine Bedeutung entnehmen – im Wohnzimmer sind es 13 °C.

Werden diese Informationen in einem bestehenden Kontext eingeordnet, der auf Erfahrungen, Wertvorstellungen und bereits erlangtes Fachwissen aufsetzt, entsteht »Wissen«. Man erkennt Muster und verarbeitet diese. Im Beispiel ist das Wohnzimmer mit »nur« 13 °C »zu kalt«.

Das hierdurch entstandene neue Wissen, aufbauend auf bereits existierendem Wissen, ermöglicht es nun, durch Anwendung ein Handeln abzuleiten. Dies setzt voraus, dass ein gewisses »Verstehen« stattgefunden hat. Regelbasierte Systeme verstehen üblicherweise nicht, sie wenden nur erlerntes Wissen an und führen darauf Anpassungsmaßnahmen durch (z. B. Temperatur im Wohnzimmer um 3 °C erhöhen).

Wissen im aktuellen Kontext einordnen zu können und daraus Prinzipien abzuleiten, setzt Verstehen voraus – einen kognitiven und analytischen Prozess. Dies ist eine Fähigkeit, die oft nur dem menschlichen Verstand zugeschrieben wird. Sie führt zu »Weisheit«. Dieser Schritt der Pyramide betrifft die Zukunft, die vorherigen Schritte dagegen die Vergangenheit.

In diesem Zusammenhang sollte auch die soziale Konstruktion von Wissen betrachtet werden.

Open Data ist in anderen Ländern bereits eine Kultur des offenen Einsehens und Teilens. Diese Kultur führt dazu, dass Nutzerinnen und Nutzer nicht jedes Mal bei null beginnen müssen. Sie können dadurch auf der nachvollziehbaren Arbeit vieler aufbauen und müssen nicht grundlegende Prinzipien selbst jedes Mal neu erarbeiten. Dies geht einher mit dem gängigen Bild des Umgangs von Wissen in der Wissenschaft – wir stehen auf den Schultern von Riesen und können mit unserer Arbeit ein Stückchen weiter sehen<sup>13</sup>.

Datenersteller Datenbereitsteller (Data Provider) und Datennutzerinnen und -nutzer (Data Consumer) sind im Bereich von Open Data häufig nicht dieselbe Organisation. Um Open Data über Grenzen von Organisationen nutzen zu können, sind offene, standardisierte, digitale und maschinenlesbare Schnittstellen für Daten und Metainformationen wie z. B. Formate und deren Semantik, Lizenzen und Bedingungen zwischen Datenbereitstellern und Datennutzerinnen und -nutzern erforderlich. Nur so können die bereitgestellten Daten von einer größtmöglichen Nutzergruppe wiederverwendet werden.

Datenersteller müssen in der Lage sein, die Relevanz und den Nutzen ihrer Daten außerhalb ihrer Organisation zu antizipieren bzw. zu verstehen. Bei der Vermittlung zwischen Bedarfsseite (Data Consumer) und Angebotsseite (Data Provider) können beispielsweise dedizierte Ansprechpartner in Organisationen, Kataloge von existierenden Daten (verfügbar über Webschnittstelle) oder auch Private-Public-Partnerships (PPP) für einen Dialog zwischen Bedarfsträgern und möglichen Anbietern von Daten unterstützen. Auch die Kombination verschiedener Datenquellen ist denkbar und kann in bestimmten Fällen dazu führen, dass so erst die vollen Potenziale von Daten ausgeschöpft werden können.

Datenbereitsteller müssen in der Lage sein, die Ziele und Bedarfe von (potenziellen) Datennutzerinnen und -nutzern hinsichtlich der Formate der und des Zugriffs auf die bereitzustellenden Daten zu verstehen und technisch umzusetzen. Als typische PPP-Formate sind hier beispielsweise Hackathons auf Basis vorhandener Open Data zu nennen. Ebenso sind Plattformen zum Abgleich von möglichen Datenangeboten in einer Art Katalog und den Datenbedarfen für die Entwicklung und den Betrieb innovativer Dienstangebote denkbar; so kann eine zielgerichtete Vermittlung und Ausrichtung von Open Data Angeboten unterstützt werden.

13 ↗ [https://de.wikipedia.org/wiki/Zwerg\\_auf\\_den\\_Schultern\\_von\\_Riesen](https://de.wikipedia.org/wiki/Zwerg_auf_den_Schultern_von_Riesen)

# 3. Open Data Fähigkeiten für Behörden und Unternehmen

## 3.1. Open Data Literacy: Die Phasen im Lebenszyklus

Wie der Prozess der Wissenserzeugung zeigt, können aus Daten Informationen entwickelt und anschließend Handlungen abgeleitet werden. Daten werden erst dann zu Informationen, wenn sie in einen Kontext gestellt oder mit anderen Informationen und Daten kombiniert werden. Aus den daraus gewonnenen Informationen leitet sich wiederum Wissen ab, das für spätere Handlungen essenziell ist.

Hierfür müssen Daten in einer verständlichen Form vorliegen und ohne bereits durchgeführte Interpretationen bereitstehen.

Um Daten zur Abbildung und Modellierung der Realität identifizieren zu können, kommen Kriterien infrage, wie z. B.:

- Welche Werte (aus Beobachtungen, Messungen oder statistische Erhebungen) von physikalischen, mathematischen oder verwaltungstechnischen Objekten sind verfügbar?
- Welche Art der Abbildung eignet sich am besten? (Als Zeichenketten oder Binärdaten (Encoding))?
- Welche Grundlagen bestehen, um aus den Daten Informationen zu extrahieren und Wissen zu erzeugen?
- Wie ist der Kontext der Erfassung bzw. Erhebung der Daten?
- Bilden die Merkmale und Aspekte ein vollständiges Bild der zu erfassenden Realität?

Für datenbasierte Entscheidungsprozesse in Behörden und Unternehmen ist ein grundlegendes Datenverständnis der Mitarbeitenden sowie ein Überblick der erhobenen Daten essenziell. Tagtäglich werden in der Verwaltung Entscheidungen getroffen. Durch die immer steigende Komplexität dieser (bspw. in den Bereichen Mobilität und Infrastruktur sowie Klima) sollten Daten die Grundlage für Entscheidungen sein. Hinter dem Begriff datenbasierte Entscheidungsfindung (Data Driven Government) versteckt sich das Treffen von Verwaltungsentscheidungen auf Grundlage von Daten und dessen Analyse. Grundlage ist somit ein funktionierendes Datenmanagement, eine Aufbereitung und Analyse sowie die Interpretation der vorliegenden Daten.

Um dies ausführen zu können, werden folgende notwendige Skills für Data Literacy betrachtet:

- Verständnis für Daten
- Wissen zu Quellsystemen für Data Collection
- Umgang mit Daten und Datenbanken
- Wissen über Data Governance und Datenmanagement
- Erfahrung mit Data Engineering, Data Enrichment und Datenfusion
- Fähigkeit der Nutzung von Data Storage und Automation
- Verarbeitung und Analyse großer und kleiner Datenmengen
- Data Science für strukturierte und unstrukturierte Daten
- Kenntnisse zur Interpretation von Daten und der Analyseergebnisse
- Fähigkeit aus Ergebnissen und Interpretationen Handlungen abzuleiten und umzusetzen

Aus der Datenkompetenz heraus ergeben sich verschiedene Phasen im Lebenszyklus von Open Data. Dabei wird auf das zugrunde liegende Modell für Datenkompetenz<sup>14</sup> zurückgegriffen, dieses um verschiedene Aspekte erweitert und für den Open Data-Kontext an relevanten Stellen angepasst. Hier werden insbesondere auch die beiden zentralen Stakeholder-Gruppen Data Provider und Data Consumer von Open Data berücksichtigt.

Der Open Data Lebenszyklus umfasst somit sechs Phasen. Diese sind:

1. Etablierung einer Datenkultur: umfasst alles, was erforderlich ist, damit eine Organisation im Denken und Handeln den Wert von Daten erkennt und offen für die eigene Nutzung oder die Nutzung durch Dritte ist.
2. Bereitstellung von Daten: umfasst die Beschreibung von Datenanwendungen, den richtigen Umgang mit Regulatorik bis hin zur qualitätsgesicherten Bereitstellung von Daten für die jeweiligen Anwendungszwecke und die dafür erforderlichen technischen Fähigkeiten.
3. Auswertung von Daten: umfasst die Analyse von Daten, die Visualisierung der Ergebnisse der Analysen und Verbalisierung derselben zur späteren Interpretation.
4. Validierung der Ergebnisse: umfasst die Fähigkeit, die richtigen Schlüsse aus den vorliegenden Auswertungen zu ziehen.
5. Validierung der Daten: umfasst die Fähigkeit, die richtigen Schlussfolgerungen der Daten auf Basis der Ergebnisse und Aspekten der Datenbeschaffung zu ziehen.
6. Etablieren von Handeln: umfasst die Fähigkeit, aus Interpretation und Ergebnissen und Daten die richtigen Handlungsmöglichkeiten und Aktionen abzuleiten und durchzuführen.

14 <https://de.wikipedia.org/wiki/Datenkompetenz>

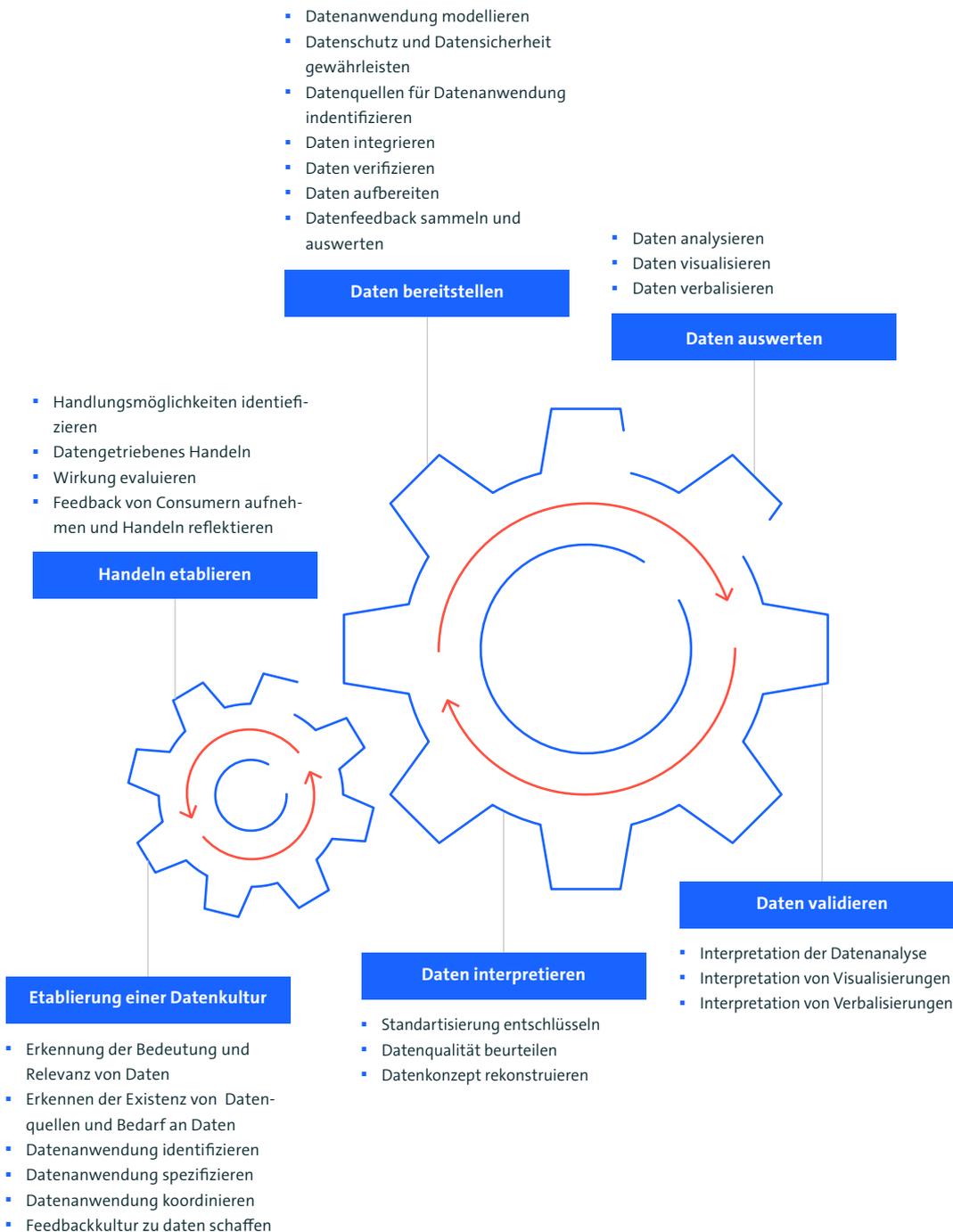


Abbildung 4: Sechs Phasen des Open Data Lebenszyklus und Fähigkeiten je Phase

In jeder der Phasen des Open Data Lebenszyklus sind bestimmte Fähigkeiten bei Mitarbeitenden der Organisation, die mit Open Data umgehen, erforderlich. Diese sind in Abbildung 4 dargestellt. Die Fähigkeiten werden im Folgenden in einer Tabelle 1 näher beschrieben. Dabei ist bei vielen Fähigkeiten entlang des Open Data Lebenszyklus ein Zusammenwirken von Data Provider und Data Consumer erforderlich, um eine praktische Nutzbarkeit der bereitgestellten oder bereitzustellenden Daten zielgerichtet zu gewährleisten.

## 3.2. Data Literacy Fähigkeiten von Data Provider und Consumer Organisationen entlang der Phasen im Open Data Lifecycle

Die in Tabelle 1 beschriebenen Fähigkeiten haben nicht den Anspruch, dass jede Person in einer Organisation, die sich mit Open Data als Provider oder Consumer befasst, diese besitzen muss. Vielmehr erfordert es dezidierte Rollen, die Data Literacy auf verschiedenen Ebenen in die Organisation und ihre Abläufe einbringen. Dabei reichen die Ebenen von der konzeptionellen und modellierenden Ebene bis hinunter auf die technisch-operative Ebene.

Tabelle 1 ist in sechs Abschnitte entsprechend der Phasen aus Abbildung 2 gegliedert. Innerhalb der Phasen werden die relevanten Fähigkeitsbereiche benannt und beschrieben und folgend in Kompetenzen und Kenntnisse der Stakeholder Data Provider und Data Consumer detailliert.

Fähigkeitsbereiche	Beschreibung	Kompetenzen und Kenntnisse
<b>Etablierung einer Datenkultur</b>		
<b>Erkennen der Bedeutung und Relevanz von Daten</b>	Das grundsätzliche Verständnis von Digitalisierung und offenen Daten (Open Data) sowie internetbasierten Angeboten und Services ist auf verschiedenen Ebenen in der Organisation verstanden und verankert; es ist z. B. durch eine entsprechende Strategie oder Leitlinie manifestiert.	Der Provider und der Consumer <ul style="list-style-type: none"> <li>▪ können Digitalstrategien, insbesondere Open Data Strategien, verstehen, entwickeln und umsetzen,</li> <li>▪ verstehen Digitalisierung, deren Anforderungen sowie Nutzen,</li> <li>▪ haben das Entwickeln von und Handeln auf Basis von Strategien in ihrer Organisation verankert</li> </ul>
<b>Erkennen der Existenz von Datenquellen und Bedarf an Daten</b>	Der Data Provider ist in der Lage, interne Daten hinsichtlich ihrer Relevanz für Open Data-Anwendungen zu bewerten. Dies kann entweder eigenständig intern geschehen oder auch im Dialog mit potenziellen Data Consumern. Der Data Consumer ist in der Lage zu beurteilen, wo die Daten, die für eine bestimmte Anwendung benötigt werden, bezogen werden können. Dazu können auch Portale wie z. B. GovData <sup>15</sup> zu Hilfe genommen werden, wenn es Daten aus Behörden und der öffentlichen Verwaltung betrifft.	Der Provider <ul style="list-style-type: none"> <li>▪ besitzt die Fähigkeit, nutzenorientiert in Anwendungsszenarien zu denken und so Relevanz und Potenzial von Daten für Open Data zu erkennen</li> <li>▪ ist in der Lage, Veranstaltungen wie z. B. Hackathons, Designathons, Data Round Tables zu planen und durchzuführen, um den Dialog mit möglichen Consumern zu fördern</li> </ul> Der Consumer <ul style="list-style-type: none"> <li>▪ besitzt Kenntnisse der Open Data Landschaft, z. B. in Behörden in Deutschland und Europa,</li> <li>▪ kennt Kataloge/Portale und Quellen zu Open Data Angeboten,</li> <li>▪ kann für seine Zwecke und Ziele passende Open Data Angebote zielgerichtet finden</li> </ul>
<b>Datenanwendung identifizieren</b>	Der Data Provider identifiziert auf Basis von Daten oder bekannten Datenbedarfen eine konkrete Aufgabenstellung, die mithilfe von Open Data gelöst werden kann. Der Provider ist in der Lage eine Vorstellung vom möglichen Wertbeitrag der Daten zu entwickeln. Der Consumer ist in der Lage eine mögliche Datenanwendung zu identifizieren und möglichen Providern verständlich darzulegen.	Der Provider <ul style="list-style-type: none"> <li>▪ kann Datenanwendungen identifizieren und beschreiben,</li> <li>▪ ist in der Lage mögliche Anwendungen auf Basis seiner Daten zumindest grob zu skizzieren,</li> <li>▪ versteht den Wertbeitrag der für die Anwendung relevanten Open Data</li> </ul> Der Consumer <ul style="list-style-type: none"> <li>▪ besitzt die Fähigkeit, Anwendungsfälle zu erfassen, zu beschreiben und an Provider verständlich zu kommunizieren</li> </ul>

Fähigkeitsbereiche	Beschreibung	Kompetenzen und Kenntnisse
<b>Datenanwendung spezifizieren</b>	Consumer und Provider definieren im Dialog miteinander minimale und optionale Anforderungen an Daten und Abgrenzungen zu anderen Aufgaben bzw. Anwendungen. Der Provider strukturiert den Prozessablauf in Objekte und deren Beziehungen, leitet messbare Objekte und Hypothesen über deren Zusammenhänge ab. Im erneuten Dialog zwischen Consumer und Provider werden weitere Anforderungen an Daten identifiziert.	<p>Der Provider</p> <ul style="list-style-type: none"> <li>kann die Daten, deren Bereitstellung inkl. erforderlicher Prozesse im Detail technisch beschreiben.</li> <li>ist in der Lage, messbare Objekte zu beschreiben und Hypothesen über deren Verhalten sowie Zusammenhänge zwischen Parametern zu verstehen</li> </ul> <p>Der Consumer</p> <ul style="list-style-type: none"> <li>besitzt die Fähigkeit aus den Anwendungsfällen exakte Datenbedarfe erkennen und zu spezifizieren,</li> <li>ist in der Lage, Anwendungsfälle und Datenbedarfe dem Provider verständlich (auch als Feedback) zu kommunizieren</li> </ul>
<b>Datenanwendung koordinieren</b>	Der Provider plant und koordiniert interne Datenprojekte, die die benötigten Daten aus den Quellen über ggf. mehrere Schritte über Schnittstellen nach außen bereitstellt. Dabei können auch externe Dienstleister oder der Consumer hinzugezogen werden. Die Planung und Koordinierung eines Datenprojekts können ggf. mit Beteiligung von weiteren Personen (aus interdisziplinären Bereichen) erfolgen.	<p>Der Provider</p> <ul style="list-style-type: none"> <li>ist in der Lage interne Datenprojekte (Bereitstellung) zu planen und im Verlauf zu steuern,</li> <li>ist in der Lage, die Infrastruktur (Hardware und Software) zur Bereitstellung der identifizierten Daten auf Basis einer Spezifikation im Rahmen eines Datenprojektes zu entwickeln bzw. die Entwicklung zu koordinieren,</li> <li>ist in der Lage nach Ende des Projektes (Daten stehen an einer Schnittstelle für den Abruf über Datenpipelines bereit) die Datenquellen, Pipelines und Schnittstellen sowie erforderliche Infrastruktur (Hardware und Software) zu betreiben und bei Bedarf zu warten</li> </ul>
<b>Feedbackkultur zu Daten schaffen</b>	Der Provider ermöglicht es Nutzerinnen und Nutzern, Feedback zu den Daten oder der Datenanwendung zu geben, z. B. über eine öffentliche Kommentarfunktion oder ein Forum für Provider-Consumer-Dialog. Consumer sind sich bewusst, dass ihr Feedback gehört und ernst genommen wird, und dass sie so konstruktiv zur Verbesserung beitragen. Dies kann der Beginn eines strukturierten Community-Managements zu Open Data sein.	<p>Der Provider</p> <ul style="list-style-type: none"> <li>ist in der Lage, mit Feedback sachlich und professionell umzugehen und den Consumern damit Wertschätzung zukommen zu lassen,</li> <li>ist in der Lage schnell und zielgerichtet auf Feedback zu reagieren und im Bedarfsfall auch deeskalierend zu wirken,</li> <li>kommuniziert auf Basis von Feedback auch abgeleitete Maßnahmen und deren Umsetzung auf dem gleichen Kanal, über den Feedback ankommt,</li> <li>schafft eine Möglichkeit für Consumer Feedback zu bereitgestellten Daten und allen ihren Aspekten, wie z. B. Qualität, Lieferfrequenz, Interoperabilität, zu geben</li> </ul> <p>Der Consumer</p> <ul style="list-style-type: none"> <li>ist fähig, fachliches und technisches Feedback (z. B. nach vorgegebener Struktur) zu Erfahrungen in der Nutzung der Daten zu geben</li> </ul>
<b>Daten bereitstellen</b>		
<b>Datenanwendung modellieren</b>	Der Provider entwickelt Modelle, die die zu vermessenden Objekte und deren beschreibende Variablen mit definierbaren Eigenschaften und deren Beziehungen abbilden. Dabei wird der Anwendungskontext berücksichtigt.	<p>Der Provider</p> <ul style="list-style-type: none"> <li>ist in der Lage, Daten zu modellieren, standardisieren und Abhängigkeiten zu beschreiben,</li> <li>ist fähig bei der Modellierung den zukünftigen Anwendungskontext mit einzubeziehen, z. B. nötige Lieferfrequenz von Daten, deren erforderliche Auflösung und Qualität</li> </ul> <p>Der Consumer</p> <ul style="list-style-type: none"> <li>ist in der Lage die Modelle des Providers zu verstehen,</li> <li>ist fähig aus den Modellen des Providers eigene Modelle für Datenanwendungen zu abzuleiten,</li> <li>kann Datenbedarfe (inkl. detaillierten Anforderungen an Qualität, Lieferfrequenz, etc.) klar an den Provider zu kommunizieren</li> </ul>

Fähigkeitsbereiche	Beschreibung	Kompetenzen und Kenntnisse
<b>Datenschutz und Datensicherheit gewährleisten</b>	<p>Der Provider beurteilt, ob die betroffenen Daten nach Prinzipien und Richtlinien sicherer und ethisch fundierter Datenverarbeitung bereitgestellt werden können.</p> <p>Er setzt diese Prinzipien und Richtlinien für die betroffenen Daten um bzw. ergreift adäquate technische und organisatorische Maßnahmen, falls keine eindeutigen Richtlinien definiert sind.</p>	<p>Der Provider</p> <ul style="list-style-type: none"> <li>▪ identifiziert die relevante Regulatorik, z. B. Datenschutz, Informationssicherheit, von denen die betroffenen Daten betroffen sind,</li> <li>▪ ist fähig Zwecke und Erlaubnistatbestände für eine Verarbeitung betroffener Daten zu identifizieren und rechtssicher zu begründen,</li> <li>▪ ist in der Lage Sicherheits- und Schutzbedarfe betroffener Daten und deren Verarbeitung zu bestimmen,</li> <li>▪ besitzt die technischen und organisatorischen Skills, um adäquate technische und organisatorische Maßnahmen zur Unterstützung der Zwecke und Erlaubnistatbestände zu implementieren</li> </ul>
<b>Datenquellen für Datenanwendung identifizieren</b>	<p>Der Provider identifiziert verschiedene gängige und neuartige Datenquellen (intern, extern) und bewertet deren Zugänglichkeit, Relevanz und Nutzbarkeit.</p>	<p>Der Provider</p> <ul style="list-style-type: none"> <li>▪ kann Datenquellen identifizieren, klassifizieren und die Nutzbarkeit im Anwendungskontext bewerten,</li> <li>▪ kann die Relevanz und damit Mehrwerte der Daten in geplanten Anwendungskontexten beurteilen,</li> <li>▪ kann Zugänglichkeit von Daten beurteilen und einschätzen, ob technische Hindernisse für die Zugänglichkeit überwunden werden können</li> </ul> <p>Der Consumer</p> <ul style="list-style-type: none"> <li>▪ kann aus seinem Anwendungskontext heraus Bedarfe an weiteren Daten und Datenquellen identifizieren und dem Provider mitteilen (Feedback)</li> </ul>
<b>Daten integrieren</b>	<p>Der Provider persistiert identifizierte Daten und stellt diese automatisiert über definierte Schnittstellen bereit. Der Consumer liest diese bereitgestellten Daten in verschiedenartigen Formaten über die Schnittstellen ein.</p>	<p>Der Provider</p> <ul style="list-style-type: none"> <li>▪ kann die definierten Schnittstellen auf Basis von persistenten Daten erstellen und bereitstellen,</li> <li>▪ ist fähig definierte Schnittstellen zu pflegen und Änderungen an Formaten und Schnittstellen im Bedarfsfall vorzunehmen und rechtzeitig an die Consumer zu kommunizieren,</li> <li>▪ kann Änderungen von Schnittstellen gegenüber Auswirkungen auf Kompatibilität beurteilen und abwägen</li> </ul> <p>Der Consumer</p> <ul style="list-style-type: none"> <li>▪ kann vom Provider bereitgestellte Schnittstellen integrieren und auf Änderungen adäquat reagieren</li> </ul>
<b>Daten verifizieren</b>	<p>Der Provider erkennt und bereinigt Inkonsistenzen in Daten, wie z. B. fehlende, verfälschte Daten oder Ausreißer, und reichert im Bedarfsfall Daten um relevante Datenpunkte an. Das Ziel ist dabei, eine hinreichend hohe Datenqualität zu erreichen. Der Provider und der Consumer filtern und verknüpfen relevante Daten für ihre spezifischen Fragestellungen.</p>	<p>Der Provider</p> <ul style="list-style-type: none"> <li>▪ kann Inkonsistenzen in Daten erkennen und bereinigen, wie z. B. Daten transformieren, standardisieren, anreichern, korrigieren, fehlende Daten rekonstruieren</li> </ul> <p>Der Provider und der Consumer:</p> <ul style="list-style-type: none"> <li>▪ können relevante Daten aus den bereitgestellten definierten Schnittstellen auslesen, filtern, verknüpfen und gemäß Anwendungszweck zu analysieren,</li> <li>▪ beherrschen verschiedene adäquate Methoden für die Bereinigung von Daten im Fall von Inkonsistenzen, die sich auf Datenqualität negativ auswirken können</li> </ul>
<b>Datenfeedback sammeln und auswerten</b>	<p>Der Provider erhebt, sammelt, strukturiert und wertet systematisch Feedback von Consumern aus.</p>	<p>Der Provider</p> <ul style="list-style-type: none"> <li>▪ ist in der Lage, Consumer-Feedback einzuholen und so Rückschlüsse auf Verbesserungsbedarfe, z. B. hinsichtlich Dateninhalten, Datenqualität, definierten Schnittstellen,</li> <li>▪ beherrscht verschiedene Techniken, um qualifiziertes Feedback von Consumern zu erhalten, z. B. Fragebogen, Expertenrunden, Nutzerforum, Beobachtung von Nutzerverhalten</li> </ul>

Fähigkeitsbereiche	Beschreibung	Kompetenzen und Kenntnisse
<b>Daten auswerten</b>		
<b>Daten analysieren</b>	Der Consumer erstellt statische und dynamische Visualisierungen unter Zuhilfenahme geeigneter Werkzeuge sach- und zweckorientiert und setzt diese ein.	<p>Der Consumer</p> <ul style="list-style-type: none"> <li>▪ versteht die Fragestellung der erforderlichen Analysen</li> <li>▪ besitzt Mitarbeitende, die Kompetenz in der Analyse von Daten besitzen, wie in <b>Tabelle 2</b> für die Rolle Data Scientist und weitere abgeleitete Rollen beschrieben</li> <li>▪ kann unter Anwendung von adäquaten Methoden (explorative, deskriptive, prädiktive Statistik, KI, Mustererkennung, etc.) die vom Provider erhaltenen Daten analysieren und in Aussagen überführen</li> </ul>
<b>Daten visualisieren</b>	Der Consumer erstellt statische und dynamische Visualisierungen unter Zuhilfenahme geeigneter Werkzeuge sach- und zweckorientiert und setzt diese ein.	<p>Der Consumer</p> <ul style="list-style-type: none"> <li>▪ kann Ergebnisse der Datenanalyse in tabellarischer, grafischer und maschinenlesbarer Form darstellen</li> <li>▪ kann diese Ergebnisse unter Anwendung geeigneter Werkzeuge für die Zwecke der Anwendung verwenden</li> </ul>
<b>Daten verbalisieren</b>	Der Consumer interpretiert die Analyseergebnisse sowie verwendete Statistiken und Modelle, prüft Interpretationen und überführt Ergebnisse in Aussagen.	<p>Der Consumer</p> <ul style="list-style-type: none"> <li>▪ kann Analyseergebnisse und Visualisierungen sowie verwendete Statistiken und Modelle sinnvoll interpretieren, Wissen extrahieren und in verbale Form umwandeln</li> <li>▪ kann vorhandene implizit und explizite Interpretationen kritisch prüfen, z. B. hinsichtlich Konsistenz, und bei Bedarf anpassen</li> </ul>
<b>Ergebnisse validieren</b>		
<b>Interpretation der Datenanalyse</b>	Der Consumer interpretiert Datenanalysen, um Schlussfolgerungen daraus zu ziehen. Dabei ist das Ziel, Handlungsempfehlungen oder Aktionen der Anwendung abzuleiten.	<ul style="list-style-type: none"> <li>▪ Der Consumer</li> <li>▪ kann Ergebnisse von Datenanalysen innerhalb der Anwendungsszenarien interpretieren</li> <li>▪ kann aus Interpretationen Schlussfolgerungen mit direktem Bezug zu seinem Anwendungszweck ziehen und Handlungsempfehlungen und Aktionen der Anwendung ableiten</li> </ul>
<b>Interpretation von Visualisierungen</b>	Der Consumer interpretiert Grafiken, die aus Daten und Datenanalysen entstanden sind und zieht Rückschlüsse auf wesentliche Elemente und Zusammenhänge.	<p>Der Consumer</p> <ul style="list-style-type: none"> <li>▪ kann Bedeutung und Maßstäbe von Skalen korrekt erkennen und damit auf Plausibilität prüfen</li> <li>▪ ist in der Lage, Fehler oder Ausfälle der Datenerhebung zu erkennen und zu kennzeichnen, um falsche Rückschlüsse vermeiden zu können</li> <li>▪ ist in der Lage auf Basis der Interpretationen und Modelle Prognosen zu erstellen</li> <li>▪ kann Korrektheit und Adäquatheit der Interpretationen gegenüber den Visualisierungen prüfen</li> </ul>
<b>Interpretation von Verbalisierungen</b>	Der Consumer interpretiert und bewertet statistische Kennzahlen und Modelle hinsichtlich ihrer Aussagen und hinterfragt diese. Dabei werden Aussagen aus den Daten auch mit den vorab festgelegten Hypothesen (vgl. Phase <i>Etablierung einer Datenkultur</i> , Schritt <i>Datenanwendung spezifizieren</i> ) abgeglichen.	<p>Der Consumer kann</p> <ul style="list-style-type: none"> <li>▪ Korrelationen verstehen und bewerten</li> <li>▪ Modelle verstehen und kann diese anpassen</li> <li>▪ die Qualität von Stichproben anhand verfügbarer statistischer Parameter einschätzen und Anpassungen an Parametern vornehmen</li> <li>▪ Verzerrungen in Daten erkennen, die auf fehlende, falsche oder zu viele Messwerte hindeuten, wie z. B. Outlier oder Missing Data, und Maßnahmen zur Behebung ableiten</li> <li>▪ die Mess- und Erhebungsmethodik den Daten gegenüberstellen</li> <li>▪ Aussagen aus den Daten hinsichtlich deren Übereinstimmung zu den Hypothesen beurteilen</li> </ul>
<b>Daten interpretieren</b>		
<b>Standardisierung entschlüsseln</b>	Der Consumer ist in der Lage, verwendete statistische Methoden und Transformation von Daten zu erkennen, einzuschätzen und interpretieren zu können. Dies kann auch zu besserer Nachvollziehbarkeit der Interpretationen beitragen.	<p>Der Consumer</p> <ul style="list-style-type: none"> <li>▪ versteht und beherrscht statistische Methoden und Transformation von Daten, kann diese erkennen, einschätzen, interpretieren und adäquat handhaben</li> </ul>

Fähigkeitsbereiche	Beschreibung	Kompetenzen und Kenntnisse
<b>Datenqualität beurteilen</b>	<p>Der Consumer besitzt die Fähigkeit, Datenanalysen und Visualisierungen sowie mitgelieferte Informationen zu Daten vor dem Hintergrund Datenqualität (Relevanz hinsichtlich des Anwendungsszenarios, Aktualität, Verlässlichkeit) zu beurteilen.</p> <p>Der Provider ist in der Lage, Anomalien in den Daten zu erkennen und den oder die Consumer zu informieren.</p>	<p>Der Consumer</p> <ul style="list-style-type: none"> <li>kann Qualität der Quellen und Übertragungswege und die Vorverarbeitung der Daten in der Bereitstellung und Beschaffung einschätzen</li> <li>beherrscht statistische und technische Methoden zur Beurteilung von Datenqualität</li> </ul> <p>Der Provider</p> <ul style="list-style-type: none"> <li>ist in der Lage, die Qualität der eigenen Daten statistisch und technisch zu bewerten</li> <li>ist fähig Consumer über Anomalien zeitnah informieren</li> </ul>
<b>Datenkonzept rekonstruieren</b>	Im Bedarfsfall ist der Consumer in der Lage, die Datenkonzeption des Providers ausreichend detailliert aus den gelieferten Daten zu rekonstruieren.	<p>Der Consumer</p> <ul style="list-style-type: none"> <li>kennt die Datengrundlage und identifiziert potenzielle Fehlschlüsse</li> <li>ist in der Lage, die Datengrundlage und -konzeption bei Bedarf aus den gelieferten Daten, Analysen und Konzepten zu rekonstruieren, falls der Provider nicht ausreichende Informationen zur Verfügung gestellt hat</li> <li>ist fähig im Bedarfsfall Feedback an den Provider zu erforderlichen Anpassungen an Konzepten, Analysen, Interpretationen und Visualisierungen zu übermitteln</li> </ul>
<b>Handeln etablieren</b>		
<b>Handlungsmöglichkeiten identifizieren</b>	Der Consumer identifiziert konkrete Handlungsmöglichkeiten aus den interpretierten, visualisierten und verbalisierten Daten. Der Consumer besitzt eine Vorstellung vom möglichen Wertbeitrag der Daten bei der Ableitung von Handlungsmöglichkeiten.	<p>Der Consumer</p> <ul style="list-style-type: none"> <li>verfügt über geeignete Mittel, um Handlungsmöglichkeiten zu identifizieren</li> <li>bezieht die relevanten Stakeholder in der Organisation in die Ableitung von Handlungsmöglichkeiten ein</li> <li>ist in der Lage auf Basis ihres Wertbeitrags verschiedene Handlungsmöglichkeiten präzise zu beschreiben und zu bewerten</li> </ul>
<b>Datengetriebenes Handeln</b>	Der Consumer integriert Ergebnisse aus Analysen in den Entscheidungsprozess und richtet sein Handeln nach diesen Ergebnissen aus.	<p>Der Consumer</p> <ul style="list-style-type: none"> <li>ist dazu fähig auch in Situationen mit zahlreichen Stakeholdern Entscheidungen systematisch herbeizuführen und zu dokumentieren</li> <li>ist in der Lage getroffene Entscheidungen in seiner Organisation konsequent umzusetzen</li> </ul>
<b>Wirkung evaluieren</b>	Der Consumer beobachtet die Wirkung der umgesetzten Handlungen und bewertet diese anhand von Kriterien.	<p>Der Consumer</p> <ul style="list-style-type: none"> <li>ist in der Lage, verschiedene Ansätze zur Bewertung der Wirkung von durchgeführten Handlungen systematisch zu durchzuführen, z. B. vorher-nachher-Vergleiche, Gruppenvergleiche.</li> <li>ist in der Lage zielgerichtet zu diesem Zweck sinnvolle und treffende Kriterien oder Kennzahlen zur Bewertung zu erheben</li> </ul>
<b>Feedback von Consumern aufnehmen und Handeln reflektieren</b>	<p>Der Provider nutzt Feedback-Kanäle, um von Consumern Rückmeldungen zu den Daten zu erhalten. Hier können verschiedene Aspekte wie z. B. Vollständigkeit, Korrektheit, Verlässlichkeit oder Nutzbarkeit relevant sein.</p> <p>Der Provider setzt auf Basis des Feedbacks Verbesserungen in den betroffenen Phasen des Data Literacy Lifecycle um.</p>	<p>Der Provider</p> <ul style="list-style-type: none"> <li>ist in der Lage Feedback zu Daten und Datenanwendung systematisch zu sammeln, zu strukturieren und auszuwerten</li> <li>ist in der Lage aus Feedback Verbesserungen abzuleiten und diese nach Dringlichkeit und weiteren Kriterien zu priorisieren und im Anschluss entlang des Data Literacy Lifecycle zielgerichtet umzusetzen</li> <li>ist in der Lage mögliche und priorisierte Verbesserungen an die Consumer transparent zu kommunizieren</li> </ul>

Tabelle 1: Fähigkeiten im Open Data Lebenszyklus

### 3.3 Data Literacy: Rollen in Open Data-Projekten und deren Kenntnisse

In Tabelle 1 wurden die erforderlichen Fähigkeiten der Provider und Consumer von Open Data zur professionellen Planung und Durchführung von Open Data-Projekten und zum späteren Betrieb von Produkten oder Services umfassend und detailliert beschrieben. Diese Fähigkeiten können konsequent in Rollen und notwendige Skills der Rollen für eine erfolgreiche Umsetzung von Open Data-Projekten heruntergebrochen und abgebildet werden.

Tabelle 2 fasst die essenziellen Rollen, zu verstehen als Aufgabengebiete in einem Open Data-Projekt, zusammen und erörtert sie jeweils in einer kurzen Beschreibung, die einen Abriss der erforderlichen Skills der Personen, die die Rollen ausfüllen, wiedergibt.

Rolle mit Bezug Data Literacy und Datenprojekten	Beschreibung
<b>Domänenexpertinnen und -experten</b> (Mitarbeitende von Fachabteilungen, z. B. Analystinnen und Analysten)	Domänenexpertinnen und -experten kennen die Problemstellungen und Bedürfnisse sowie deren Kontext und die fachlich relevanten Prozesse und Datenquellen. Zu Beginn von Projekten arbeiten sie konzeptionell und gegen Ende für die Validierung der entwickelten Lösungen. Typischerweise sind sie in relevanten Fachabteilungen tätig, wie beispielsweise Marketing, Fertigung, Vertrieb, Accounting oder Supply Chain.
<b>Data Scientist</b>	<p>Data Scientists verbinden die Sichtweise von Fachabteilung und Management. Sie erkennen so Bedürfnisse, die sich aus der Problemstellung und Konzeption ableiten und sind in der Lage davon zu abstrahieren. Dabei soll der Data Scientist den Nutzen in den verfügbaren Daten erkennen können, um so Entscheidungsfindungsprozesse im Unternehmen unterstützen zu können. Data Scientists sind in der Lage, verschiedenste Datenquellen zusammenzuführen, aufzubereiten, zu bereinigen, zu analysieren und zu visualisieren.</p> <p>In komplexeren Projekt-Setting ist oft eine Unterteilung der Rolle Data Scientist in die Bereiche</p> <ul style="list-style-type: none"> <li>▪ Machine Learning (ML) Engineer - Spezialisierung auf Training, Optimierung und Validierung von ML-Modellen</li> <li>▪ Data Engineer - Bereitstellung von benötigter IT-Infrastruktur für Datenverarbeitung (falls keine explizite Rolle Data Engineer im Projekt vorhanden ist, übernimmt die Rolle Data Scientist diese Aufgaben weitestgehend)</li> <li>▪ Data Modeler - Modellierung von Daten und Zusammenhängen einzelner Attribute, sodass Daten nutzbar, auswertbar, interpretierbar werden</li> </ul> <p>Daneben besitzen Data Scientists querschnittlich Skill im Bereich Software-Entwicklung, um Datenverarbeitung in Projekten möglichst effizient zu automatisieren.</p>

Rolle mit Bezug Data Literacy und Datenprojekten	Beschreibung
<b>Data Engineer</b>	<p>Die Rolle des Data Engineer ist in Datenprojekte auf Softwareentwicklung spezialisiert. Dabei kommen Data Engineers im Schwerpunkt in frühen und späten Phasen eines Projektes zum Einsatz.</p> <p>In frühen Phasen gehören zu den primären Aufgaben das</p> <ul style="list-style-type: none"> <li>▪ Schaffen von Schnittstellen zwischen beteiligten Systemen (Systemen, die Datenquellen sind, und Systeme, die Daten zusammenführen und Daten-Analysen ausführen</li> <li>▪ Aufsetzen von Datenbanken, Data Warehouses oder Cloud-Instanzen und weiterer erforderlicher IT-Infrastruktur</li> <li>▪ Beschaffen, Aufbereiten und Anreichern von Daten</li> </ul> <p>In späten Phasen gehören zu den primären Aufgaben das</p> <ul style="list-style-type: none"> <li>▪ Bereitstellen der Ergebnisse von Analysen</li> <li>▪ Integrieren der Ergebnisse in den operativen Alltag</li> </ul> <p>Eine besondere Ausprägung dieser Rolle findet sich in der Rolle des <b>Citizen Data</b> Engineer, welche Konzepte und Lösungen des Data Engineers operativ in die Breite trägt und dabei operativ Unterstützung von nicht-technischen, aber geschultem Personal erhalten kann. So wird die Brücke zwischen Data Science und Umsetzung mittels Informationstechnik gebildet.</p>
<b>Data &amp; Analytics Architect</b>	<p>Die Rolle Data &amp; Analytics Architect zeichnet als fachliche Führungsrolle für Planung und Aufbau der technischen Infrastruktur verantwortlich. Dabei sind neben funktionalen Aspekten wie z. B. fachlichen Themen zu Daten, Analysen und Speicherung auch nicht-funktionale Aspekte wie z. B. Sicherheit, Verfügbarkeit, Compliance zu Regulatorik wie z. B. Datenschutz und ethische Verantwortung für die Verarbeitung von Daten relevant.</p> <p>Data &amp; Analytics Architects entwickeln Data-Governance Richtlinien und setzen sie um und tragen dafür Sorge, dass die Datenstrategie mit der Unternehmensstrategie einhergeht.</p> <p>Dabei erhalten sie operativ Unterstützung von sogenannten <b>Data Stewards</b>, die Daten und Anfragen zu Daten auf Regelkonformität prüfen und die Einhaltung der erforderlichen Qualitätsstandards sicherstellen.</p>
<b>Data Product Manager / Owner</b>	<p>Data Product Manager bzw. Owner verantworten den Aufbau von Datenprodukten und -services entlang deren Lebenszyklus. Sie entwerfen Vision, Strategie und Roadmap für das Daten-basierte Produkt bzw. den Daten-basierten Service. Dafür identifizieren sie die relevanten Business-Cases, holen Bedürfnisse und Anforderungen der Kundschaft und Nutzenden ein und entwerfen auf dieser Basis Konzepte und Kriterien für Erfolgsmessung der Datenprojekte, den sie auch mittels der Kriterien bewerten und an das obere Management berichten.</p>

Tabelle 2: Open Data-bezogene Rollen in Provider und Consumer Organisationen

In der Umsetzung von Datenprojekten können grundsätzliche noch weitere Rollen mit spezifischen Fähigkeiten erforderlich sein, wie z. B. Software Engineer, Projektmanager, Data Steward oder Citizen Data Engineer. Diese werden hier nicht im Detail erläutert, da sie außerhalb des thematischen Kerns »Data Literacy« liegen. Eine personelle Ausstattung von Open Data-Projekten mit den in der Tabelle 2 beschriebenen Rollen ist wesentlicher Erfolgsfaktor für die Planung und Entwicklung von solchen Projekten und den späteren erfolgreichen Betrieb eines so entwickelten Datenprodukts oder -services.

Abbildung 5 ordnet die oben beschriebenen Rollen den Phasen des Open Data Lifecycle zu. So wird anschaulich klar, welche Rolle in welcher Phase besonders wichtig ist.

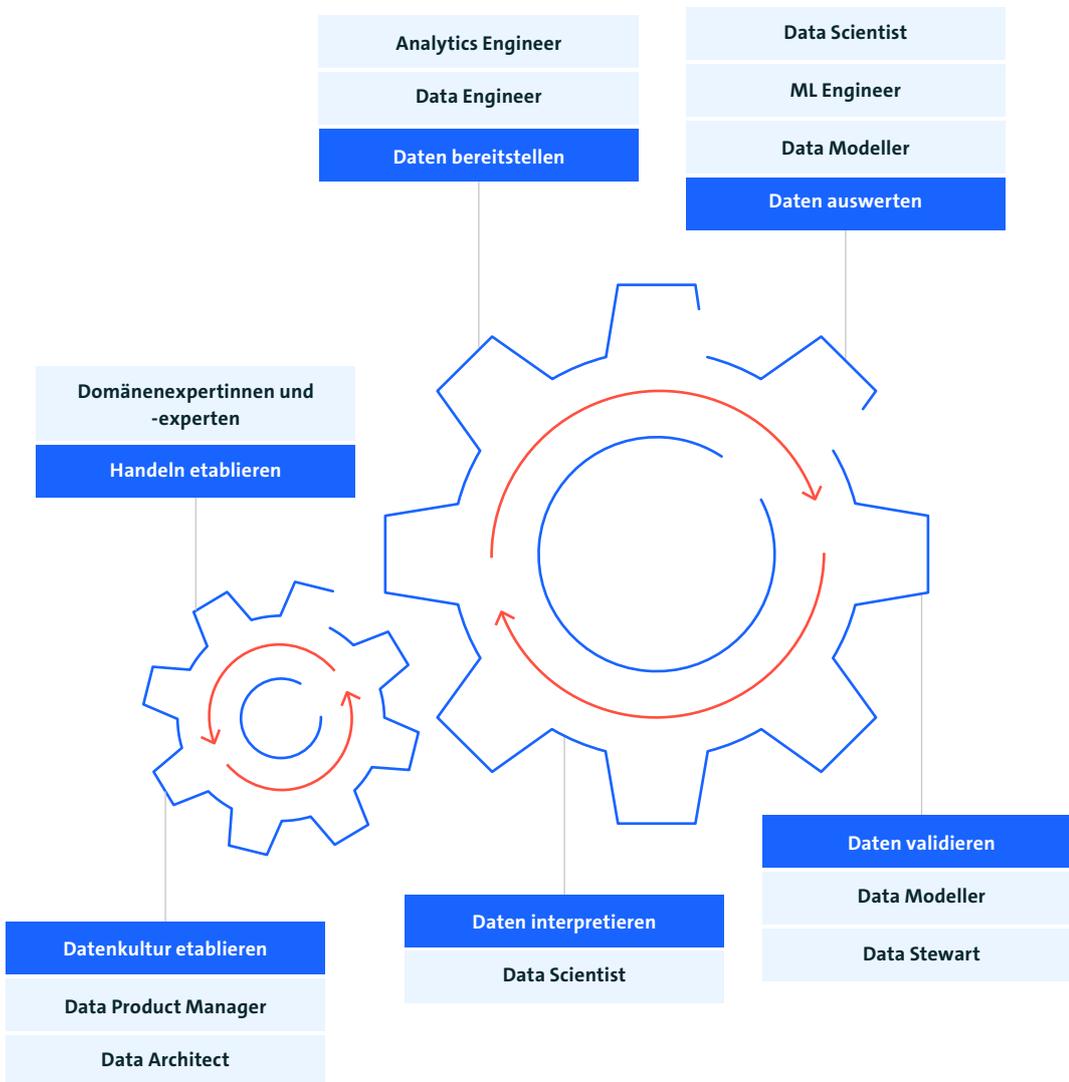


Abbildung 5: Welche Rolle ist in welcher Phase aktiv?

# 4. Szenarien für Open Data mit professioneller Data Literacy

## 4.1 Szenario 1: Parkplatzdaten auf Autobahnparkplätzen

Die LKW-Parkplatzproblematik speziell an Autobahnen ist allgemein bekannt und nimmt zu. Nach Angaben von Kravag und des Bundesverbandes Güterkraftverkehr Logistik und Entsorgung (BGL) fehlen derzeit bundesweit rund 40.000 LKW-Parkplätze<sup>16</sup>. Eine Suche nach vorhandenen Plätzen unter der Einhaltung der Lenkzeiten ist daher oft mit Zeit und Stress verbunden. Dies führt zu einem erhöhten Risiko (Falschparker) sowie unnötigem CO<sub>2</sub> Ausstoß. Das alles macht den LKW-Fahrer Job noch unattraktiver, was den Fachkräftemangel in dieser Branche verstärkt.

Eine erhebliche Hilfe bei der Parkplatzsuche können digitale Lösungen bieten. Diese könnten die Auslastung der Parkplätze in Echtzeit ermitteln und prognostizieren. Als Input für solche digitalen Lösungen können unterschiedliche Daten verwendet werden, z. B. Telematik Daten der Flottenmanagementsysteme<sup>17</sup>, Parkplatzsensorik für einzelne Parkplätze oder auch LiDAR-Sensoren<sup>18</sup>, die im Gegensatz zu Videokameras eine hochauflösende 3-D-Abbildung der Raststätte und gleichzeitig Datenschutzkonformität gewährleisten können.

Die beschriebenen Daten können durch Datenerzeuger selbst (Raststättenbetreiber) sowie über Datendienstleister (Telematik-Anbieter) bereitgestellt und genutzt werden. Dabei sollten die Open-Data-Prinzipien, wie Zugänglichkeit und Maschinenlesbarkeit berücksichtigt werden. Nach Empfehlung des Deutschen Datenportals<sup>19</sup>, können die eingerichteten Benutzer- und Programmierschnittstellen (API) den Zugang zu Daten erleichtern. Auch die Verwendung von etablierten Datenformaten, wie CSV oder JSON im Gegensatz zu PDF-Dokumenten sollte bevorzugt werden. Unter anderem spielt die Aktualität der Daten für das Szenario eine entscheidende Rolle, was gleichzeitig auch die größte Herausforderung darstellt.

Bei der Nutzung kann man zwischen Primär- und Sekundärnutzung unterscheiden. Zur Primärnutzung gehören vor allem Verkehrsmanagement und lenkung.

Diese bringen eine ganze Reihe an Vorteilen:

- Eine Möglichkeit der flächendeckenden Parkplatzsuche für das Transportgewerbe kann zu einer Entlastung für die LKW-Fahrer führen und dadurch die Tätigkeit attraktiver machen.

<sup>16</sup> ↗ [https://www.kravag-truck-parking.de/wp-content/uploads/2021/04/KRAVAG\\_eWhitepaper\\_final.pdf](https://www.kravag-truck-parking.de/wp-content/uploads/2021/04/KRAVAG_eWhitepaper_final.pdf)

<sup>17</sup> ↗ <https://www.bmdv.bund.de/SharedDocs/DE/Artikel/DG/mfund-projekte/solp.html>

<sup>18</sup> ↗ <https://de.wikipedia.org/wiki/Lidar>

<sup>19</sup> ↗ [https://www.govdata.de/documents/10156/18448/GovData\\_Open-Data-Kriterien\\_der\\_Sunlight\\_Foundation.pdf/dca8fea0-8e04-4de0-8531-2bc3e8d4abc0](https://www.govdata.de/documents/10156/18448/GovData_Open-Data-Kriterien_der_Sunlight_Foundation.pdf/dca8fea0-8e04-4de0-8531-2bc3e8d4abc0)

- Informationen über die vorhandenen Plätze sowie deren Auslastung kann für die Planung der Infrastrukturoptimierung bzw. Ausweitung durch Bund und Länder genutzt werden (effektiver Ausbau von Parkplätzen).
- Neben der Entlastung von LKW-Fahrern unterstützt die Lösung für die Parkplatzsuche ganz allgemein die Verkehrssicherheit auf der Autobahn.

Unter Sekundärnutzung kann die Parkplatzbelegung u. a. für Auslastungsprognosen für die Tank- und Ladeinfrastruktur oder das Gaststättengewerbe genutzt werden.

Die bessere Verteilung der Parkenden trägt auch zum Klimaschutz bei, da der Parksuchverkehr, der bisher ca. drei Tonnen CO<sub>2</sub> pro Lkw und Jahr ausstößt<sup>20</sup>. Die Bereitstellung von Daten auf Open-Data-Portalen wie der Mobilithek, ermöglicht auch die Nutzung z. B. durch Verkehrsmanagement-Zentralen und die Wissenschaft.

Betriebswirtschaftliche Ziele, wie die Minimierung administrativer Aufwände für Speditionen und Fahrpersonal sind ebenso realisierbar. Die Verwendung von Telematikdaten führt im Gegensatz zur ortsfesten Ausrüstung von Lkw-Parkplätzen (Sensorik wie Kameras, Zählschleifen, etc.) zu signifikanter Kostenreduktion.

Da zurzeit keine flächendeckende einheitliche Lösung existiert, können erst durch eine Datenfusion (das Zusammenführen von Daten aus verschiedenen Quellen) neue Nutzungsperspektiven entstehen.

Anforderungen an die Datenqualität, wie die Konsistenz, Vollständigkeit und Nachvollziehbarkeit, sollen unter der Berücksichtigung der Anonymisierung und Nichtrückverfolgbarkeit realisiert werden.

## 4.2. Szenario 2: Abwasser-Monitoring als Frühwarnsystem bei drohenden Corona-Ausbrüchen

Die Corona-Pandemie hat das deutsche Gesundheitssystem an seine Belastungsgrenzen gebracht und Mängel bei der Digitalisierung von Meldekettens aufzeigt. Das hat im Rückblick Innovationsprozesse angestoßen, die uns in der Zukunft helfen werden, mit vergleichbaren Situationen umzugehen.

Als Beispiel seien Pilotprojekte zum Echtzeit-Monitoring von Corona-Virenbelastungen im Abwasser genannt, die zur Prävention und als Frühwarnsystem in einer Pandemie geeignet sind – insbesondere bei verminderten Testzahlen oder aufgrund asymptotischer Verläufe.<sup>21</sup> Studien haben gezeigt, dass die über das Abwasser-Monitoring ermittelten Infektionszahlen den Erhebungen von Testergebnissen über die Gesundheitsämter um bis zu zehn Tage voraus sind.<sup>22</sup>

20 ↗ <https://www.eurotransport.de/artikel/umfrage-zur-parkplatznot-wie-gross-ist-das-lkw-parkplatz-problem-6646255.html>

21 ↗ <https://www.rki.de/DE/Content/Institut/OrgEinheiten/Abt3/FG32/Abwassersurveillance/Abwassersurveillance.html>

22 ↗ <https://www.bundestag.de/resource/blob/848440/484417db3669d04d1980cadb380de056/Abwasser-Covid-19-data.pdf>

Ziel des Abwasser-Monitorings ist es, einen guten Überblick über das tatsächliche Infektionsgeschehen zu erhalten. Steigende Zahlen können dabei helfen, eine Verschärfung oder Rückkehr des Virus in einem Gebiet frühzeitig zu erkennen und Maßnahmen zum Schutz der Menschen (wieder) aufzunehmen bzw. bei sinkenden Zahlen diese Maßnahmen aufzuheben. Sie erlauben darüber hinaus, Trends in einer Pandemie-Entwicklung frühzeitig zu erkennen und die Wirksamkeit von getroffenen Maßnahmen zu überwachen.<sup>23</sup>

Beim Abwasser-Monitoring werden zwei Datenströme analysiert. Zum einen die Konzentrationen von SARS-CoV-2-Viren in Abwasser-Mischproben. Die Proben werden von den kommunalen Abwasserbetrieben am Kläranlagenzulauf bzw. im Kanalisationsnetz entnommen. Die im Labor ermittelte Viruslast wird auf Basis der an das Kanalisationssystem angeschlossenen Bevölkerungszahl und der Abwassermenge normalisiert. Zum anderen wird die über die Gesundheitsämter erfasste Anzahl von (PCR-getesteten) Infizierten je Tag im zugehörigen Abwasserentsorgungsgebiet erfasst. Letzteres stellt den Zusammenhang her zwischen den Abwassermesswerten und dem über Testungen bekannten Infektionsgeschehen innerhalb eines Gebietes. Die abstrakten Konzentrationswerte werden so in interpretierbare Informationen ähnlich dem Inzidenz-Wert überführt. Viele Faktoren beeinflussen die Werte, wie z. B. ein zeitlicher Versatz zwischen Messung und Tests, das Verhältnis von Getesteten zur Dunkelziffer nicht-getesteter Infizierter, aber auch nicht erfasste Infektionen im Abwasser aufgrund nicht ganztägiger Mischproben.

Die Erfassung und Weitergabe aller Daten sollte digital erfolgen: Die Abwasserproben können über digitale Formulare erfasst werden. Die erfassten Daten sollten direkt in eine zentrale Cloud-Infrastruktur fließen, in der alle Beteiligten zusammenarbeiten können.

Die Corona-Testdaten kommen aus den Meldesystemen der Gesundheitsämter über eine API. In der zentralen Infrastruktur werden sie mit den Abwasserentsorgungsgebieten verschnitten und so anonymisiert.

Basierend auf den zentral zur Verfügung stehenden Daten können dann medienbruchfreie, nutzergruppenspezifische Sichten in Form von Karten und Dashboards für verschiedene Stakeholder entstehen – für politische Entscheider, Gesundheitsämter oder die Öffentlichkeit.

23 <https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32021H0472&from=EN>

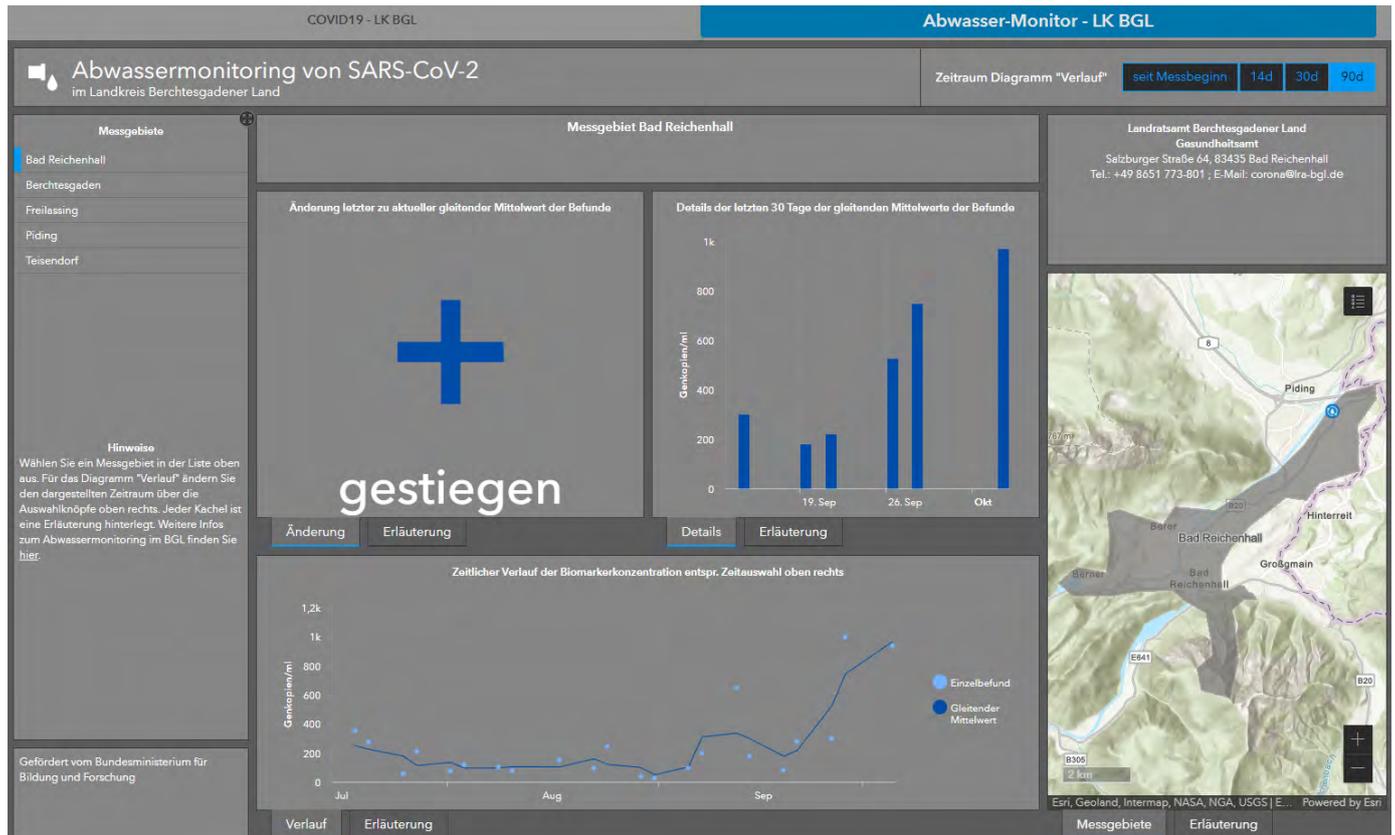


Abbildung 6: Dashboard zum Abwasser-Monitoring von SARS-CoV-2 im Landkreis Berchtesgadener Land [Stand: April 2023]<sup>24</sup>

Auf einem Abwassermonitoring-basierende Systeme bieten ein kosteneffizientes Instrument<sup>25</sup> zur Vorsorge und nachhaltigen Steuerung von Reaktionsmaßnahmen – nicht nur bei SARS-CoV-2. Sie können genauso bei der künftigen Verbreitung anderer Krankheitserreger oder von Schadstoffen zum Einsatz kommen.<sup>26</sup>

Im beschriebenen Szenario kommen Daten aus verschiedenen Quellen zusammen: Messdaten, Meldedaten, Daten zu Abwasserentsorgungsgebieten, statistische Kennzahlen. Einige, aber nicht alle Daten sind Open Data. Die Verschneidung und Normalisierung anonymisiert sie und erzeugt einen mit der Inzidenz vergleichbaren Kennwert, der verständlich und als Open Data geeignet ist. Der Verlauf und Trend eines solchen Kennwerts kann politische Entscheidungsprozesse unterstützen. Das Monitoring des Kennwerts kann während einer Pandemie als Basis oder Teil von Informationsprodukten wie Karten oder Dashboards – maßgeschneidert für bestimmte Zielgruppen – für Transparenz und demokratische Teilhabe sorgen.

24 <sup>↗</sup> <https://experience.arcgis.com/experience/4bc6e16430264215ba5ad1dd14ffbc78/page/Abwasser-Monitor---LK-BGL/>

25 <sup>↗</sup> [https://ec.europa.eu/environment/water/water-urbanwaste/info/index\\_en.htm](https://ec.europa.eu/environment/water/water-urbanwaste/info/index_en.htm)

26 <sup>↗</sup> <https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32021H0472&from=EN>

# 5. Zusammenfassung

Der Leitfaden »Open Data Literacy« richtet sich an Organisationen aus Wirtschaft und Verwaltung, die Daten als offene Daten bereitstellen oder nutzen möchten. Data Literacy ist, so zeigt es auch der Leitfaden aus verschiedenen Perspektiven, integraler Bestandteil – ja sogar eine Grundvoraussetzung – um erfolgreich mit offenen Daten arbeiten zu können.

Der Leitfaden stellt dabei wesentlich einen Open Data Lifecycle vor, der spezifisch für die Dualität von Provider und Consumer von Open Data sowie öffentliche wie auch private Organisationen zugeschnitten ist. Der Lifecycle besteht aus sechs Hauptphasen, die je von verschiedenen Einzelfähigkeiten gestützt werden. Dabei ist es – je nach Situation und Bedarfslage für Open Data – nicht zwingend erforderlich, jede einzelne Fähigkeit jeder Lifecycle Phase exakt nach Muster umzusetzen. Es darf und soll pragmatisch ein Customizing der Fähigkeiten stattfinden, so dass ein praktikabler Ansatz eines Open Data Lifecycle in öffentlichen und privaten Organisationen entsteht, der auch zur vorliegenden Situation passt.

Die Rollen aus dem Data Science Bereich unterstützen dabei sowohl Provider als auch Consumer bei der Auswahl von passenden Mitarbeiterinnen und Mitarbeitern für Open Data Projekte sowie bei der Ausschreibung von Stellen und der Auswahl von Personen für Neueinstellungen.

Neben den fachlichen Themen wie z. B. dem Ablauf von der Identifikation über die Bereitstellung bis hin zur Nutzung von Open Data in digitalen Services sowie den dazu erforderlichen Fähigkeiten der Organisationen und Personen, werden auch eine Reihe von Beispielen für Open Data aus der Praxis sowie praxisnahe Szenarien mit hohem Nutzenpotenzial für die Gesellschaft dargestellt. Sie erklären letztlich pragmatisch, wie nützlich Open Data sein können und sein werden, wenn das Thema Open Data umfassend und zielgerichtet vorangetrieben und umgesetzt wird.

Abschließend sei erwähnt, dass Open Data immer zwei Seiten benötigt: den Provider und den Consumer. Somit gilt auch hier das Eingangsmotto dieses Leitfadens:

»Erfolg entsteht nicht allein durch die individuelle Anstrengung, sondern durch die Zusammenarbeit und das Engagement eines Teams« – also auch das »Team« aus Open Data Provider und Consumer.

Bitkom vertritt mehr als 2.200 Mitgliedsunternehmen aus der digitalen Wirtschaft. Sie erzielen allein mit IT- und Telekommunikationsleistungen jährlich Umsätze von 190 Milliarden Euro, darunter Exporte in Höhe von 50 Milliarden Euro. Die Bitkom-Mitglieder beschäftigen in Deutschland mehr als 2 Millionen Mitarbeiterinnen und Mitarbeiter. Zu den Mitgliedern zählen mehr als 1.000 Mittelständler, über 500 Startups und nahezu alle Global Player. Sie bieten Software, IT-Services, Telekommunikations- oder Internetdienste an, stellen Geräte und Bauteile her, sind im Bereich der digitalen Medien tätig oder in anderer Weise Teil der digitalen Wirtschaft. 80 Prozent der Unternehmen haben ihren Hauptsitz in Deutschland, jeweils 8 Prozent kommen aus Europa und den USA, 4 Prozent aus anderen Regionen. Bitkom fördert und treibt die digitale Transformation der deutschen Wirtschaft und setzt sich für eine breite gesellschaftliche Teilhabe an den digitalen Entwicklungen ein. Ziel ist es, Deutschland zu einem weltweit führenden Digitalstandort zu machen.

**Bitkom e.V.**

Albrechtstraße 10  
10117 Berlin  
T 030 27576-0  
[bitkom@bitkom.org](mailto:bitkom@bitkom.org)

[bitkom.org](https://www.bitkom.org)

**bitkom**