

# Doch was sind Daten? Bedeutungsdimensionen eines vielschichtigen Begriffs

EVA JABLONKA, BERLIN & KATHARINA JABLONKA-COHEN, BERLIN

**Zusammenfassung:** Im Fokus des Beitrags stehen Überlegungen zum Begriff der Daten, der im Diskurs zur „data literacy“ als verbindendes Element erscheint. Zunächst setzt sich der Beitrag mit dem Datenbegriff in konzeptionellen Überlegungen zur „data literacy“ an Hand ausgewählter Beiträge einander, die aus unterschiedlichen Fachkulturen hervorgehen. Daran anschließend werden Bedeutungen des Datenbegriffs durch eine begriffs geschichtliche Annäherung aufgespürt. Der Beitrag schließt mit einem Ausblick auf Möglichkeiten, wie unterschiedliche Begriffsdimensionen und epistemische Konnotationen an Hand von Beispielen thematisiert werden können.

**Abstract:** The article focuses on the notion of data, which appears as a unifying concept in the discourse on 'data literacy'. The article begins by examining the concept in selected contributions on 'data literacy' that emerge from different specialist cultures. Subsequently, different meanings of the term 'data' are traced through a genealogical approach. The article concludes with an outlook on how different conceptual dimensions and epistemic connotations can be addressed on the basis of examples.

## 1. Einleitung

Im Schulfach Mathematik fällt der Umgang mit „Daten“ traditionell in die Domäne des Stochastikunterrichts und des mathematischen Modellierens, untermauert mit konzeptionellen Überlegungen zur *statistical literacy* und *mathematical literacy*. Nun werden mit *data literacy* in Konzeptionen zur Bildung in der digitalen Welt – unter Verweis auf die Relevanz von aus Daten abgeleiteten Entscheidungen auf Basis von „Big Data“ und Algorithmen – auch Themen des Informatikunterrichts, der Gesellschaftswissenschaften und der Philosophie in den Blick genommen (Gould, 2017; Gould, 2021; Heinemann et al., 2018; Wild, 2017; Wolff et al., 2017). Auch in Empfehlungen öffentlicher Akteure wird die Bedeutung von „Datenkompetenz“ betont, um „Fähigkeiten für ein selbstbestimmtes Handeln im digitalen Zeitalter“ (BRG, 2021, S. 41 ff.) auszubilden und den Herausforderungen „einer rasanten Transformation der digitalen Informationsökosysteme“ gerecht zu werden (KomZS, 2024, S. 6). Mitunter finden sich sogar Formulierungen, die Daten als das

Öl des 21. Jahrhunderts bezeichnen und *data literacy* innerhalb einer Gesellschaft als notwendig für die Abwendung von potenziellen Bedrohungen verstehen (Ridsdale et al. 2015; Schüller 2019). Für den Mathematikunterricht fordern dementsprechend Bildungsstandards und Rahmenlehrpläne einen Beitrag „zur Vermittlung von Kenntnissen sowie der Entwicklung von Fertigkeiten und Fähigkeiten für die Orientierung in der durch Digitalisierung geprägten Gesellschaft“ (KMK, 2023, S. 24), der den kompetenten „Umgang mit Daten“, oder die „kritische Reflexion und die Einordnung von Informationen aus verschiedenen digitalen Darstellungen“ (SenBJF Berlin, 2023, S. 6) einschließt.

Der vielschichtige Begriff der „Daten“ entpuppt sich nun in diesem Fächergrenzen überschreitenden Diskurs als verbindendes Element, aber auch als ein potentieller fachsprachlicher Fauxami. Denn der Datenbegriff hat in der Mathematik und Statistik, in den empirischen Natur- und Sozialwissenschaften und in der Informatik unterschiedliche Bedeutungen und verschiedene epistemische Konnotationen, die sich im Alltagsgebrauch und im öffentlichen Diskurs nicht immer klar unterscheiden lassen.

Im pädagogisch-didaktischen Diskurs erweist sich dieser Umstand als problematisch, wenn in konzeptionellen Überlegungen zur *data literacy* der Datenbegriff unterbeschrieben bleibt, verengt wird oder gar nicht erläutert wird, wie auch Emmenegger (2018) bereits feststellte:

Daten werden reifiziert und selbst bei Spezifizierungen mehrheitlich informationstheoretisch dimensioniert; eine solide Einordnung von Konzepten wie Fakten, Wissen, Tatsachen, Informationen bleibt aus (Emmenegger, 2018, S. 238).

Daher sehen wir den dringenden Bedarf einer genaueren Betrachtung des allgegenwärtigen Begriffs der „Daten“. Dazu versuchen wir zunächst, Bedeutungsdimensionen des Datenbegriffs in konzeptionellen Überlegungen zur Datenkompetenz oder *data literacy* aufzuspüren, die aus unterschiedlichen Fachkulturen hervorgehen. Wir gehen dabei nicht davon aus, eine gemeinsame Kernbedeutung zu finden, sondern suchen nach expliziten Bestimmungen, Abgrenzungen oder Einordnungen. In einem zweiten Schritt versuchen wir, die in den konzeptionellen Überlegungen sichtbar gewordenen verschiedenen begrifflichen Nuancen und Kon-

notationen von „Daten“ zu präzisieren und die Leerstellen zu füllen. Dazu gehen wir begriffsgeschichtlich vor, um Ausdifferenzierungen oder Bedeutungsverschiebungen, die der Datenbegriff in verschiedenen Praktiken der Verwendung von „Daten“ erfährt, präzisieren zu können.

## 2. Datenbegriff in schulbezogenen konzeptionellen Überlegungen zu *data literacy*

Die folgende Übersicht stellt den Datenbegriff in Beiträgen zu *data literacy* aus verschiedenen fachlichen Perspektiven dar, mit dem Ziel, die erwähnten unterschiedlichen Bedeutungsdimensionen des Begriffs im pädagogischen Diskurs zu verdeutlichen. Der Fokus für die Auswahl liegt dabei auf schulbezogenen Beiträgen, die konzeptionellen Überlegungen zu *data literacy* nachgehen.

Engel (1999) spricht aus der Perspektive der Didaktik der Statistik bzw. *statistical literacy*, angelehnt an Biehler und Weber (1995), von der Notwendigkeit einer *data literacy*, die er mit dem Begriff der Datenkompetenz für den deutschsprachigen Diskurs greifbar macht.<sup>1</sup> Diese Überlegungen sind Resultat seiner frühen Beobachtung, dass „eine Überflutung mit Tabellen, Graphiken und Statistiken“ (Engel, 1999, S. 77) im öffentlichen Diskurs stattfindet, deren Grund er in der fortschreitenden Mathematisierung der empirischen Wissenschaften mit der gleichzeitig durch neue Technologien ermöglichten Sammlung und Verarbeitung von „Massendaten“ sieht, die mit der zunehmenden Bedeutung von Daten in politischen und ökonomischen Zusammenhängen einhergeht. Vom „mündigen Bürger“ werde für die Ausbildung von Datenkompetenz „eine Fähigkeit zum Einschätzen, Bewerten und Argumentieren mit *quantitativ dargebotenem Material* erwartet, auf [die] unsere allgemeinen Bildungseinrichtungen vorbereiten müssen.“ (Engel, 1999, S. 77, Hervorhebung hinzugefügt). Für die Didaktik der Statistik ergibt sich daraus, dass neben „Fähigkeiten zum Darstellen, Zusammenfassen und Analysieren gegebener Daten und grundlegenden Kenntnissen über die Mathematik des Zufalls auch ein Können und Wissen um Probleme der Datenerhebung und Konzepte des Stichprobendesigns“ wichtig seien, sowie eine Intuition dafür, „welche Information in Daten stecken kann sowie die Fähigkeit, zufallsbedingte Variabilität in Daten zu erkennen und bei allen Zufallseinflüssen wesentliche Informationen aus den Daten zu extrahieren (Engel, 1998)“ (Engel, 1999, S. 78). Hier geht es also um die Betrachtung quantifizierter Daten mit einem Fokus auf ihre statistischen Eigenschaften. Der Datenbe-

griff entspricht am ehesten dem der quantitativen empirischen Sozialwissenschaften, den Engel von einem Datenbegriff im „Schutzraum idealisierter Umgebungen und abstrakter Zufallsmodelle“ (Engel, 1999, S. 79) abgrenzt, der in Curricula mit einem Fokus auf mathematische Grundlagen der Statistik vorherrsche. Diese Kenntnisse seien nicht hinreichend für das Lernziel der Datenkompetenz, für die eine Beschäftigung mit Daten, die realen Situationen oder Problemen entspringen, wichtig sei. In der international diskutierten Konzeption der „Civic Statistics“ ist ein solcher Datenbegriff innerhalb der *statistical literacy* ebenfalls im Bildungskontext wiederzufinden (vgl. Gal, 2022a, Gal et al., 2022b), wobei Gal et al. (2022b) *data literacy* als eine sich mit „Civic Statistics“ überlappende Konzeption verstehen (Gal et al., 2022b, S. 84). Der Datenbegriff wird darin nicht explizit angesprochen, sondern anhand der Nutzung und Entstehung von Daten in Kontexten von aktueller politischer und sozialer Relevanz verdeutlicht, wobei Parallelen zu Gal's (2002) Ausführungen zur *statistical literacy* zu finden sind.

Weiterhin wird *data literacy* in verschiedenen Veröffentlichungen, teils aus unterschiedlichen Fachrichtungen, mit dem Feld der Data Science in Verbindung gebracht (z. B. Gould, 2024; Heinemann et al., 2018; Ludwig & Thiemann, 2020; Schüller et al., 2019). Auch Informatik wird explizit oder implizit durch die Betonung von Digitalität, maschinellen Verarbeitungsmethoden oder durch Schlagwörter wie ‚Big Data‘ erwähnt. Infolgedessen fließen in diese curricularen Überlegungen zu *data literacy* Datenbegriffe der Informatik und Data Science ein, welche den Fokus auf das digitale Format von Daten legen. Dies soll im Folgenden anhand der Beiträge von Heinemann et al. (2018), Ridsdale (2019) und Schüller et al. (2021) beispielhaft dargestellt werden.

Heinemann et al. (2018) unternehmen den Versuch, ein Data Science Curriculum für die Sekundarstufe in Deutschland zu skizzieren, wobei Data Science dabei als „the art of generating information and knowledge from data in various forms“ (Heinemann et al., 2018, S. 1) verstanden wird, welche insbesondere auf Methoden der Informatik und Statistik beruht. *Data literacy* wird darin ähnlich wie bei Ridsdale et al. (2015) als eine Reihe von Kompetenzen verstanden, die bei der Beschäftigung mit Daten relevant werden. Zugleich weisen Heinemann et al. (2018) auf die Gefahr der Schwammigkeit einer solchen Definition hin und verfolgen deshalb den Anspruch, diese Kompetenzen innerhalb

eines Data Science Prozessmodells zu konkretisieren. Für die Ausformulierung eines Data Science Curriculums sowie eines ersten Entwurfs von Unterrichtmodulen, welche *data literacy* Kompetenzen einbeziehen, begeben sie sich explizit auf die Suche nach einem geeigneten Datenbegriff und beschreiben vier verschiedene Ansätze: Erstens den Datenbegriff aus den Bildungsstandards der Gesellschaft für Informatik e.V. (GI), der laut Heinemann et al. (2018) im deutschen Informatikunterricht etabliert ist. Dieser wird jedoch aufgrund seiner eingeschränkten Anwendbarkeit<sup>2</sup> kritisiert. Zweitens nennen die Autor:innen den Datenbegriff der sogenannten „DIKW-Pyramide“ (Ackoff, 1989). Der darin enthaltene formale Datenbegriff stammt aus den Informationswissenschaften und sieht Daten als jeder Erkenntnisfindung vorgängig an (Rowley, 2007). Drittens diskutieren Heinemann et al. (2018) den Datenbegriff der Statistikdidaktik und zitieren Cobb & Moore (1997): „Data are not just numbers, they are numbers with a context“ (Heinemann et al., 2018, o. S.). Der vierte Datenbegriff wird am Diskurs um ‚Big Data‘ festgemacht und anhand einer gängigen Charakterisierung mit Bezug auf Geschwindigkeit, Wahrhaftigkeit, Vielfalt und Volumen beschrieben. Obwohl Heinemann et al. (2018) die Frage um einen geeigneten Datenbegriff zum Schluss offenlassen, solle der Datenbegriff der GI als Ausgangspunkt genutzt werden, um eine möglichst hohe Akzeptanz in der deutschen Informatikdidaktik zu sichern.

Ridsdale et al. (2015) beschäftigen sich in ihrer Literaturanalyse zwar nicht mit einem explizit schulbezogenen Ansatz zu *data literacy*, dennoch ist deren Arbeit für diesen Kontext relevant, da sich deren Definition von *data literacy*<sup>3</sup> in schulbezogenen konzeptionellen Ansätzen wiederfindet und damit auch der Datenbegriff im Zusammenhang der *data literacy* übernommen wird (z. B. Grillenberger, 2019; Grillenberger & Romeike, 2018; Schüller et al., 2019). Die Notwendigkeit von *data literacy* wird bei Ridsdale et al. (2015) insbesondere anhand von gesellschaftlichen ökonomischen Herausforderungen und auf individueller Ebene mit Verweis auf persönliche Entscheidungsfindung verdeutlicht. Eine Explikation des Datenbegriffs ist nicht zu finden, aber implizit geht aus den Ausführungen hervor, dass vor allem digitalisierte Daten gemeint sind. Gleichzeitig wird darauf hingewiesen, dass auch die Fähigkeit, gute und schlechte Repräsentationen zu unterscheiden, zentral sei, wofür freilich der Bedeutungsgehalt der Daten relevant wird. Eine Definition, die das digitale Format ebenfalls als

entscheidend darstellt, lässt sich auch bei Schüller et al. (2021) finden. Die Autor:innen definieren „Daten“ im Hinblick auf die Datenstrategie der Bundesregierung in einer für Bildungsprozesse angedachten Data-Literacy-Charta wie folgt: „Daten sind digitale Abbilder von realen Phänomenen, Gegenständen und Prozessen – das gilt für alle Anwendungsfelder“ (S. 3). Hier zeigen sich die Schwierigkeiten einer expliziten Bestimmung des Datenbegriffs, indem zwar einerseits die Form fokussiert wird, aber andererseits die inhaltliche Bedeutung mitgemeint ist.

Aus medienpädagogischer Perspektive warnt Gebre (2022, S. 1081) in seiner Literaturanalyse, *data literacy* sei ein potentielles „buzzword“ und kritisiert eine zu enge kompetenzorientierte Fassung, die der Alltagserfahrung von Lernenden im Kontext von im Überfluss vorhandenen, öffentlich zugänglichen Daten in verschiedenen Formaten nicht gerecht werde und so deren zukünftige Möglichkeiten eher einschränke als erweitere. Dementsprechend weit gefasst fällt sein Datenbegriff aus, der nicht nur quantitative, sondern auch qualitative (z. B. Textnachrichten) oder „visuelle“ Daten einschließt (z. B. Bilder), die auf unterschiedliche Weise analysierbar und verwendbar seien (Gebre, 2018, S. 333). Aus medienpädagogischer Perspektive steht naturgemäß im Fokus, in welchem „Medium“ sich das Rohmaterial der Daten zeigt, die Kategorisierung ist allerdings theoretisch unterbestimmt. Sie verweist aber auf semiotische Dimensionen und verlässt den engen Fokus auf quantifizierte oder digitalisierte Daten.

Auch Emmenegger (2018) befasst sich mit *data literacy* aus einer (medien-)pädagogischen Perspektive und problematisiert zunächst die „Überbewertung von Datenkompetenz“ sowie die zu kurz greifenden „Grundannahmen zu Daten“ (Emmenegger, 2018, S. 233) und benennt dabei mehrere Aspekte. Zum einen führt er den „Herstellungscharakter“ von Daten an, welcher im Schatten der Fokussierung auf Nutzung und Anwendung von Daten verloren ginge. Zudem benennt er die im Diskurs angenommene Faktizität sowie das uneindeutige Format von Daten, welches scheinbar zwischen expliziter Digitalität und möglicher Digitalität schwankt und dennoch als definierende Eigenschaft herangezogen wird, wie auch wir im Hinblick auf die oben erwähnten Beiträge festgestellt haben. Weiterhin nimmt Emmenegger (2018, S. 236 f.) kritisch auf die sogenannte DIKW-Pyramide Bezug, die auch Heinemann et al. (2018) als ungeeignete Basis für konzeptionelle Überlegungen ansehen. Aufbauend auf

diesen Überlegungen, umschreibt Emmenegger einen Datenbegriff, welcher aus bildungstheoretischer Perspektive gedacht wird:

[...] Daten [sind] nicht nur als Ressource in der Umwelt des Menschen zu verorten, sondern es stellen sich Fragen dahingehend, wie durch eine Orientierung an Daten bzw. Datafizierungsprozessen Selbst- und Weltverhältnisse von Menschen bestimmt werden [...] beispielsweise, in welchem Zusammenhang technologische und über Medien vermittelte Weltzugänge mit dem Selbstverständnis von Mensch und Gesellschaft stehen (Emmenegger, 2018, S. 243).

Anhand von Gebre (2022) und Emmenegger (2018) wird deutlich, dass die dargestellten (medien-)pädagogischen Ausführungen zum Datenbegriff weit über die zuvor vorgestellten hinausgehen, eine präzise Bestimmung des Begriffs aber (zwangsläufig) ausbleibt, da verschiedene Fachdisziplinen naturgemäß unterschiedliche Bedeutungsdimensionen fokussieren.

Eine Kernbedeutung von „Daten“ lässt sich also im Diskurs um *data literacy* oder Datenkompetenz nicht ausmachen. Als gemeinsamer Ausgangspunkt der konzeptionellen Überlegungen erscheint die öffentliche Verfügbarkeit von (im Überfluss) vorhandenen Daten und deren Bedeutung für politische Entscheidungsprozesse und ökonomische Verwertbarkeit. Damit sind vor allem quantifizierte und digitalisierte Daten angesprochen. Wichtige Unterschiede und Leerstellen finden sich im Hinblick auf die Funktion von Daten in Erkenntnisprozessen, insbesondere hinsichtlich der Idee ihrer unmittelbaren Gegebenheit oder ihres Herstellungscharakters. So werden Daten als Zufallsvariablen bestimmt oder erscheinen als Abbilder realer Situationen, sind als Träger digitalisierter Information konzipiert oder werden als jeder Erkenntnis vorgängig betrachtet, erscheinen erst mit Blick auf bestimmte Zwecke hergestellt oder als unmittelbares sicheres Fundament.

### 3. Klärungen

Um die in den konzeptionellen Überlegungen sichtbar gewordenen verschiedenen begrifflichen Nuancen und Konnotationen von „Daten“ zu präzisieren und die Leerstellen zu füllen, greifen wir im Folgenden auf einzelne ausgewählte begriffsgeschichtlich orientierte Beiträge zurück, die wir durch eigene Recherchen ergänzen. Unser Interesse richtet sich dabei nicht auf unterschiedliche technische Begriffsverwendungen, sondern auf die den Daten zugesprochene Funktion in Erkenntnis- und Argumentationszusammenhängen und auf die Annah-

men über deren mehr oder weniger unmittelbare Gegebenheit.<sup>4</sup> Die begriffsgeschichtliche Skizze zeigt, dass der Begriff Metamorphosen hinsichtlich seiner epistemischen Konnotationen durchmacht und in verschiedenen Praktiken der Verwendung von Daten spezifische Ausdifferenzierungen erfährt, wobei historische Bedeutungsdimensionen nicht vollkommen zum Verschwinden kommen.

#### 3.1 Daten als Bestimmungsstücke einer logisch-argumentativen / kalkülmäßigen Ableitung

Für „Daten“ ist die Begriffsgeschichte an eine Wortgeschichte geknüpft, da offenbar kein alternativer Ausdruck mit einem ähnlichen Bedeutungshorizont benutzt wurde. Im Lateinischen bezeichnet „data“ (substantiviertes Partizip Plur. von latein. „dare“, geben) bekanntlich das Gegebene. Der Begriff eines Gegebenen als Ausgangspunkt für Ableitungen lässt sich bis zu den Ansätzen einer Methodenlehre der griechischen Mathematik zurückverfolgen. Diese steht wiederum in direktem Zusammenhang mit Überlegungen zur Erkenntnisgewinnung in Mathematik, Theologie und Philosophie des 16. und 17. Jahrhunderts. Das Gegebene bzw. das in einer Argumentation als unumstritten Zugegebene muss im Argument oder Kalkül zu dem Gesuchten (latein. „quaestio“) in Zusammenhang gebracht werden. Neue Erkenntnisse sollen auf diese Weise zugleich mit ihrer Darstellung begründet werden (Arndt, 1971). In den Schriften Euklids (Euclid 1625/1896), deren Verbreitung im 17. Jahrhundert eine Hochphase erlebte, bezieht sich „data“ (griech. Δεδομένα, vergleichbare Bedeutung wie im Lateinischen) auf bestimmte gleichartige mathematische Größen (Längen, Proportionen), von denen man ausgehen kann oder die man ansetzt, wenn man ein mathematisches Problem löst (Euclid 1625/1896).

Ein „datum“ als das (Zu-)Gegebene hat in dieser Verwendung (noch) keine Konnotation einer faktischen Gewissheit, da man auch von kontrafaktischen oder theologischen Annahmen ausgehen kann (Rosenberg, 2014). Ein „factum“ (substantiviertes Partizip Sing. des lateinischen Verbs „facere“, tun, machen, herstellen) ist hingegen etwas, was tatsächlich getan worden oder geschehen ist (Rosenberg, 2014).

Die von Euklid aufgezeigten Ableitungsmöglichkeiten wurden zwar in der Rezeption für entsprechende geometrische Berechnungen aus konkreten Größenangaben aufgegriffen (Sidoli, 2018). Aber auch bei Descartes werden in den berühmten „regulae ad directionem ingenii“ die als „data“ gege-

benen Bestimmungsstücke zur Wahrheitsbestimmung der in einer „quaestio“ vorliegenden Aussage nicht eingeschränkt auf Daten in Form von Zahlen oder Größen. Seine Methode bezieht sich gleichermaßen auf philosophische Fragen. Das Gesuchte wird mit dem Gegebenen bzw. bereits Zugegebenen als logisch äquivalent erkannt (Arndt, 1971).

Eine arithmetische Umdeutung rückt dann diese Verwendungsweise des „Gegebenen“ in die Nähe der Bedeutung von Daten als Berechnungsgrundlage, wenn man aus gegebenen Zahlenwerten mit Hilfe eines Kalküls zu Ergebnissen kommen möchte. Diese „Daten“ können allerdings absichtsvoll so gewählt sein, dass sie die naturgesetzliche Glaubwürdigkeit untergraben, was in historischen Rechenbüchern der frühen Neuzeit der Profilierung der spezifisch mathematischen Zugangsweise diente (Feistner, 2018).<sup>5</sup>

In den Bedeutungen ‚gegebene Größe, Angabe, Beleg‘ wird „data“ (als lateinische Pluralform) dann in den deutschen wissenschaftlichen Sprachgebrauch des 17. und vermehrt des 18. Jahrhunderts aufgenommen und im 19. Jahrhundert der eingedeutschte Plural „Daten“ benutzt (DWDS, 2024). Hier lassen sich sowohl die Assoziationen der quantifizierten Information („Größe, Angabe“) als auch die der logisch-argumentativen Voraussetzung („Beleg“) ausmachen. Rosenberg (2014) identifiziert in dem lateinisch gebrauchten Begriff eines „datum“ in der Theologie, Philosophie, Mathematik und auch der Naturphilosophie im 17. Jahrhundert noch einen gemeinsamen Kern, der auf nach allgemeiner Übereinkunft Vorausgehendes verweist.

### **3.2 Daten als umfassende und objektive Be standsaufnahme**

In der ersten Hälfte des 19. Jahrhunderts beobachtet Hacking (1982) in Europa eine „Lawine an gedruckten Zahlen“, die er mit dem Anstieg der Printmedien in Verbindung bringt. Es zeigte sich ein „fetischistische[s] Sammeln“ (Hacking, 1982, S. 280), insbesondere von Daten zur Bevölkerung. Das Preußische Statistische Bureau dient in diesem Zusammenhang als erstes Beispiel der Etablierung einer offiziellen staatlichen Statistik, die auch die Herausgabe von statistischen Informationen an Presse und Öffentlichkeit förderte (Hacking, 1982, S. 379). Diese demonstrierten die genaue Kenntnis des Zustandes und der Entwicklungen im Reichsgebiet als Grundlage einer Beurteilung der Lage und zur Auswertung von Maßnahmen.<sup>6</sup>

Der Enthusiasmus für beschreibende Statistik in der Bürokratie ging in Deutschland einher mit einer Zurückhaltung gegenüber der Anwendung probabilistischer Ideen in den Sozialwissenschaften.<sup>7</sup> Da es nicht um die Identifikation von Trends oder die Erkenntnis allgemeiner Gesetzmäßigkeiten ging, verband sich mit dem Gebrauch dieser empirischen Daten kein epistemischer Gewinn (Königl. Preuß. Statist. Bureau, 1860/61, S. 263). Vereinzelte Ableitungen aus den Daten sind meist als einfache „Resultate“ oder „Ergebnisse“ angesprochen. Der Begriff der „Fakten“ wird sehr selten benutzt, und zwar nur dann, wenn „Daten“ auf etwas Singuläres verweisen oder auf Merkwürdiges hindeuten. Die Hauptsorge galt der Sicherstellung einer zentral geregelten Datensammlung zur Herstellung von Vergleichbarkeit durch Qualitätsnormen. Der Datenbegriff suggeriert demnach Objektivität der Erhebung und Aufzeichnung sowie Genauigkeit.

### **3.3 Daten als mit Unsicherheit behaftete empirische Grundlage der Erkenntnisgewinnung**

Nachdem sich der wissenschaftliche Kontext des Begriffs „data“ vom Gebrauch in der Mathematik, Philosophie und Theologie im Laufe des 18. Jahrhunderts hin zu Teilgebieten der Geografie, Medizin, Finanzwelt und Naturgeschichte verschoben hatte (Rosenberg, 2014)<sup>8</sup>, änderte sich die Qualität naturwissenschaftlicher Daten durch die Entwicklung immer präziserer Messinstrumente. Damit einher ging die die Entwicklung von Methoden zur Schätzung regelmäßiger Messfehler<sup>9</sup>, gestützt durch ein verbessertes Verständnis möglicher Ausgänge von Experimenten und ein hypothesesstendes Vorgehen (Swijtink, 1987). Indem die Rolle persönlicher Urteile und Interpretationen bei der Bestimmung des Beobachteten verringert wurde, erlangten Daten objektiven Statuts. Sie haben epistemischen Wert für die Ableitung oder Widerlegung von Annahmen, sind aber selbst mit einer prinzipiellen Unsicherheit behaftet.

### **3.4 Daten als informationsübertragender Code**

Zu den oben dargestellten Bedeutungsdimensionen gesellt sich mit der Verbreitung von Computern ein formaler Begriff der Informationstheorie, der nun alles, was in Form eines Codes zwischen verschiedenen Computern (wie z. B. Satelliten, Mobiltelefonen oder Servern) übertragen wird, zu „Daten“ werden lässt (Aw, 2014). Diese Verwendung des Datenbegriffs in Anwendungen der „Informationstechnologie“ oder Informatik geht zurück auf die als kanonisch angesehenen Arbeiten zu einer „mathe-

matischen Theorie der Kommunikation“ von Shannon und Weaver.

Der Konzeption liegt ein abstraktes Verständnis von Kommunikation zugrunde, die Weaver (1949) nicht nur auf Sprache, sondern auch auf Musik jeglicher Art und bewegte Bilder bezieht, sowie eine formale Interpretation von Information.<sup>10</sup> Es finden sich darin die Unterscheidungen zwischen einer zu übertragenden Nachricht, die von einer Informationsquelle ausgeht, einem Übertragungsapparat, der diese in Signale bzw. Zeichen umwandelt und über einen Kanal an einen Empfänger sendet, der die Signale wieder zurück in die Nachricht verwandelt, sowie einer technischen Störung, die dabei auftreten kann. Zwar haben Nachrichten eine Bedeutung, diese sei aber für das Übertragungsproblem aus ingenieurwissenschaftlicher Perspektive nicht relevant. Zur Immunität der Theorie gegenüber inhaltlicher Bedeutung äußert Weaver unter anderem:

An engineering communication theory is just like a very proper and discreet girl accepting your telegram. She pays no attention to the meaning, whether it be sad, or joyous, or embarrassing (Weaver 1949/1964, S. 27)

Die Entstehung des Begriffs der Daten im Sinne digital übertragbarer Entitäten ist in diesem kommunikationstheoretischen Zusammenhang zu verstehen. Das Interesse richtet sich auf das Verhältnis von Daten und Rauschen, das Minimieren von Störungen oder auch auf prinzipielle unvermeidbare oder absichtliche Verzerrungen.

„Daten“ werden hier also nicht als epistemisch oder zeitlich vorgängig betrachtet, denn „Information“ gibt es bereits als Bedingung vor (und nach) der Übersetzung dieser in jene. Hervorzuheben ist zudem, dass diese Begriffsverwendung auf die Form und Materialität der Daten abzielt und nichts mit deren empirischem Bezug zu tun hat. Damit verlagert sich freilich die Frage nach der Bedeutung von Daten für Erkenntnisprozesse auf den Begriff der Nachricht bzw. auf die darin enthaltene wissenswerte Information.<sup>11</sup>

### 3.5 Daten als intransparenter Gegenstand einer Data Science

Wie erwähnt, finden sich in konzeptionellen Überlegungen zu *data literacy* mit Beteiligung der Mathematikdidaktik vermehrt Bezüge zur Data Science. Unter Beteiligung von Informatik, Statistik und anderen für die Erzeugung, Verarbeitung, Speicherung und Auswertung von „Daten“ relevanten Disziplinen versteht sich diese als neues interdiszipli-

linäres Feld, das beansprucht, auch die Perspektiven der unterschiedlichen Anwendungsfelder einzubeziehen, aus denen die maschinell zu verarbeitenden „Daten“ stammen. Oft wird jedoch bei einer Charakterisierung der Data Science der Fokus auf Techniken gelegt. Genannt werden dabei bestimmte Arten der Datengeneration (z. B. Sensorsysteme, Satellitendaten, Click-Verhalten auf Webseiten und andere „digitale Verhaltensspuren“), die Untersuchung „unstrukturierter“ Daten (z. B. Sprache, Bilder), die Erweiterung des Methodenrepertoires der Statistik sowie die Entwicklung von „subsymbolischen“ Algorithmen des „maschinellen Lernens“, die mit Beispieldaten oder Trainingsumwelten arbeiten (etwa Gafny & Ben-Zvi, 2024; Hrycej et al., 2023; Zhang & Xiong, 2022). Aus philosophischer Perspektive merken Desai et al. (2022) an, dass Data Science als wissenschaftliche Disziplin viele offene Fragen ihrer epistemologischen Fundierung aufwirft und ein prinzipielles Problem epistemischer Intransparenz mit Bezug auf Blackbox-Modelle mit sich bringt (vgl. Jablonka, 2017), welches nicht durch ein Verständnis der Techniken behoben werden kann.

### 4. Einsichten und Ausblicke

Die begriffsgeschichtliche Skizze eröffnet den Blick auf Kontinuitäten und Diskontinuitäten verschiedener Dimensionen und Konnotationen des Datenbegriffs. Der Begriff der Daten als nicht weiter hinterfragbare Bestimmungsstücke eines Beweises oder einer Argumentation deutet auf die Vorgängigkeit der Daten im Erkenntnisprozess, die sich sowohl als logisch-argumentative Voraussetzungen oder als mit Hilfe eines Kalküls verrechenbare Größenangaben verstehen lassen, unabhängig von ihrem Wirklichkeitsbezug. Die Konnotation der Objektivität und anzustrebenden Genauigkeit von Daten mit Wirklichkeitsbezug haben wir in der Verwaltungstastistik mit ihren standardisierten Methoden der Erhebung und Aufzeichnung ausgemacht, wobei dort die Bedeutungsdimension der Daten als Ausgangspunkt für verallgemeinerbare Erkenntnisse fehlt. Da die Daten zunächst nicht in Hinsicht auf eine bestimmte Auswertung angesammelt werden, erscheinen sie dort zudem als Ausdruck unmittelbarer Gegebenheiten. Die Konnotation der Objektivität im Verein mit einem epistemischen Wert für die Ableitung oder Widerlegung von Annahmen – im Rahmen eines hypothesegeleiteten Vorgehens – zeigt sich im Datenbegriff der aufkommenden empirischen Wissenschaften. Als Ergebnisse von methodisch kontrollierten Messprozessen oder Expe-

rimenten sind diese Daten im Hinblick auf Theorien nicht vorgängig und auch nicht unmittelbar und zugleich mit einer zufallsbedingten und prinzipiellen epistemischen Unsicherheit behaftet. Der Datenbegriff der Informatik fokussiert hingegen das Medium, Format und die Verarbeitungsmethoden. In seiner ursprünglichen kommunikationstheoretischen Fassung sind Daten nichts Vorgängiges, da der Informationsgehalt in der zu übermittelnden Nachricht schon vorausgesetzt wird. Der Fokus auf die Methoden der Verarbeitung der Daten zeigt sich noch stärker in der Data Science, die aber wiederum den epistemischen Wert von Daten hervorhebt, wobei der epistemische Gewinn auf intransparente Weise zustande kommt. Insbesondere durch automatisierte Techniken der Generation von Daten tritt außerdem deren Herstellungscharakter in den Hintergrund und es entsteht die Konnotation einer faktischen Gewissheit.

Wie im zweiten Abschnitt aufgezeigt, finden sich einige dieser Begriffsbestimmungen und Konnotationen in konzeptionellen Überlegungen zu *data literacy* angedeutet. So knüpft zum Beispiel der Datenbegriff der Empfehlungen der GI explizit an den der Informationstheorie an, indem Daten als „eine Darstellung von Information in formalisierter Art, geeignet zur Kommunikation, Interpretation und Verarbeitung“ gefasst werden, die „wieder zu Information [werden], wenn sie in einem Bedeutungskontext interpretiert werden“ (Röhner et al., 2016, S. 9). Es wird deutlich, dass dieser Datenbegriff nicht unabhängig vom Informationsbegriff verstanden werden kann, der nun aber von Weavers Fassung abweicht, denn Information ist hier „der kontextbezogene Bedeutungsgehalt einer Aussage, Beschreibung, Anweisung, Mitteilung oder Nachricht“ (Röhner et al., 2016, S. 9). Die kommunikationstheoretische Fundierung ist noch sichtbar, weil verschiedene illokative Funktionen von Äußerungen angesprochen sind. Zur epistemischen Funktion von Daten in verschiedenen Fachkulturen verhält sich dieser Datenbegriff neutral. Für eine Konzeption von *data literacy* erweist er sich als zu eng, worauf auch Heinemann et al. (2018) aufmerksam machen. Andere explizite Definitionsversuche führen zu einseitigen Verzerrungen des Begriffs im Hinblick auf den epistemischen Wert von Daten, beispielsweise wenn Versionen der sogenannten „DIKW“-Pyramide herangezogen werden, die – wie bereits von Emmenegger (2018) aufgezeigt – auf einem philosophisch nicht mehr haltbaren logisch-empirischen Operationalismus basiert.

Demnach bewegen sich (fächerüberschreitende) Konzeptionen von *data literacy* offenbar in einem Spannungsverhältnis zwischen Unterbestimmtheit des zugrunde gelegten Datenbegriffs oder dem Versuch einer expliziten Begriffsbestimmung, die zu Verengungen oder erkenntnistheoretischen Verzerrungen führt. Die Aufmerksamkeit für verschiedene Bedeutungsdimensionen und unterschiedliche epistemische Konnotationen, die aus der begriffs geschichtlichen Skizze hervorgehen, kann jedoch den Blick für verschiedene, auch widersprüchliche, Begriffsbestimmungen schärfen, wie wir im Folgenden andeuten.

Tatsächlich finden sich im Alltagssprachgebrauch „Daten“ relativ häufig im Zusammenhang mit „Information“ (DWDS, 2024).<sup>12</sup> Daten erscheinen in diesen Verbindungen allerdings als Vorgängiges, denn der kommunikationstheoretische Hintergrund ist nicht mehr sichtbar. Bei den Adjektivattributen sind „personenbezogen“, „gespeichert“ und „sensibel“ typisch. Diese beziehen sich auf ethische und rechtliche Aspekte vor allem mit Bezug auf digitale Daten. Ebenso findet sich im alltäglichen Sprachgebrauch die Bedeutungsdimension von „Daten“ als nicht weiter hinterfragbares Bestimmungsstück einer Argumentation oder eines Kalküls. So zeigt das Wortprofil als typische Verbindung mit einem Substantiv „Daten und Fakten“ insbesondere in Kombination mit „belastbar“. Die Verbindungen mit „Fakten“ und „belastbar“ verweisen deutlich auf epistemische Konnotationen<sup>13</sup>. Belastbarkeit suggeriert Unanfechtbarkeit und Fakten verweisen auf unmittelbare nicht anzuzweifelnde Tatsachen.

Die Idee der erschöpfenden Sammlung von ‚objektiven‘ administrativen Daten zu verschiedenen Themen in der Verwaltungsstatistik, die sich von ihrem Entstehungszusammenhang ablösen und zur Weiterverarbeitung zur Verfügung stehen, lässt sich bspw. auch in Open Government Data (OGD) ausmachen. Nun können solche Datenkorpora zwar versuchen, die Informationen einer ganzen Domäne zu erschöpfen, bleiben aber immer von dem Datenformat oder der Technologie abhängig, mit der sie generiert wurden. Spätestens bei der Nutzung durch andere, werden sie unter neue begriffliche oder theoretische Vorstellungen subsumiert. Aus der Perspektive der Mathematikdidaktik machen bereits Hancock et al. (1992) mit Bezug auf diesen Punkt darauf aufmerksam, dass geeignete Daten immer in Antizipation bestimmter Analysen, die vorgenommen werden sollen, generiert werden und dass die benutzten Instrumente und Formate der Aufzeichnung, Messung und Dokumentation

die Exploration strukturieren und zur Objektivierung der Methoden beitragen. Was dann die „Daten“ sind, ist in dieser Perspektive also wesentlich durch ihre Form mitbestimmt.

Die Konnotation der Unsicherheit lässt sich im alltäglichen Sprachgebrauch nicht ausmachen, denn Daten und Fakten werden, wie erwähnt, oft gleichgesetzt. Während aber unsichere oder fehlerbehaftete Daten immer noch „Daten“ bleiben, lässt sich das von Fakten nicht sagen. Der Begriff von Daten als mit Unsicherheit behaftete empirische Grundlage der Erkenntnisgewinnung ist eine der fundamentalen Ideen der Statistik (beispielsweise bei Burrill und Biehler, 2011). Für diesen mit probabilistischem Modellieren zusammenhängenden Datenbegriff ist zentral, dass Verteilungen von Messwerten im Vergleich mit hypothetisch angenommenen theoretischen Verteilungen betrachtet werden. Aber auch mit Bezug auf den Mathematikunterricht beobachten Burrill und Biehler (2011), dass nur Standardgrößen und diese meist ohne Berücksichtigung von Fehlern vorkommen und die Messung kategorischer Attribute kaum berücksichtigt werde. Diese Beobachtung halten wir nach wie vor für relevant, vor allem auch mit Bezug zur Darstellung von Daten und Auswertungen im öffentlichen Diskurs.

Während also dem Datenbegriff des Diskurses innerhalb der empirischen Wissenschaften immer epistemische Unsicherheit und prinzipielle Korrigierbarkeit anhaftet, konnotieren quantitative Daten im öffentlichen Diskurs Gewissheit, aus denen die Alternativlosigkeit evidenzbasierter Entscheidungen abgeleitet wird. Jablonka (2024a) vermutet, dass dies dem Bedürfnis entspringt, politische Handlungsmaßnahmen als Ergebnis rationaler Entscheidungen darzustellen, die auf gesicherten Wahrheiten beruhen, welche innerhalb einer als unkontrovers gedachten Wissenschaft erzeugt werden. Dazu finden sich zahlreiche bereits aufgearbeitete Beispiele der COVID-19 Pandemie. Ein Beispiel aus dem Klimadiskurs liefert eine graphische Darstellung, die dem Kongress der Vereinigten Staaten von einem Atmosphärenwissenschaftler und Klimatologen vorgelegt wurde, um Diskrepanzen zwischen den Projektionen der Klimamodelle und den Beobachtungsdaten aufzuzeigen. Die Darstellung sollte „zeigen“, dass diese Modelle die globale Erwärmung im Allgemeinen überbewerteten (Christy, 2016). Eine entsprechende Grafik wurde durch den Ausschluss von Unsicherheitsbereichen und Mittelwertbildungen erstellt und verändert

damit die Konnotation der Klimamodelle in Richtung epistemischer Gewissheit.

Vor allem als Ergebnisse nicht sichtbarer Datengeneration (durch Sensoren, GPS, Benutzertransaktionen auf Webservern, etc.) nehmen digitalisierte Daten, die sich wie von selbst erschaffen, ein materialisiertes Eigenleben an, das sie als unmittelbare faktische Gegebenheiten erscheinen lässt. Um den Herstellungscharakter solcher Daten, die automatisch (und ggf. unwissentlich) erzeugt und oft nur in bereits umgewandelter Form als graphische Darstellung sichtbar werden, im Unterricht nachvollziehbar zu machen, eignet sich beispielsweise die Untersuchung von Self-tracking Daten in Fitness-Apps auf Mobiltelefonen (z. B. Lee et al., 2021).

Verschiebungen der epistemischen Konnotationen im Prozess der Erzeugung, Verarbeitung und Interpretation von Daten lassen sich bspw. an interaktiven Visualisierungen aus wissenschaftlichen Untersuchungen zeigen, die manchmal in Form von anprechenden „Bildern“ die Zugänglichkeit von großen Datensätzen erhöhen.<sup>14</sup> Im Hinblick auf klassische ‚wissenschaftliche Bilder‘ macht bereits Heintz (2007) darauf aufmerksam, dass deren suggestive Faktizität die dahinter liegenden Entscheidungen für bestimmte Aufzeichnungs- und Bearbeitungsschritte der Daten verdeckt. Jablonka (2024a) ergänzt, dass aufgrund der Funktion vieler interaktiver Visualisierungen als konzeptionelle Blackboxes, die Nachvollziehbarkeit der Datentransformationen seitens der Benutzer:innen auf Grenzen stößt, was jedoch die kritische Auseinandersetzung mit deren Gebrauch als Ausgangspunkt in Argumentationszusammenhängen nicht behindert.

Die hier angedeuteten Beispiele sollen erkennen lassen, dass „Daten“ aus verschiedenen Perspektiven beleuchtet werden können, die unterschiedliche fachliche Konventionen oder verschiedene Verwendungspraktiken fokussieren. So können die damit einhergehenden Bedeutungsdimensionen und Verschiebungen der epistemischen Konnotationen in den Blick genommen werden. Eine Beschäftigung mit „Daten“ zur Herausbildung von *data literacy* könnte auf diesem Wege zu einem nuancierten Blick auf die Begriffsverwendungen im alltäglichen Sprachgebrauch und in den Medien beitragen. Das wäre eine wichtige Voraussetzung zur Entwicklung von *data literacy*, die nach unserer Auffassung über eine enge kompetenzorientierte Fassung hinausgeht und unter anderem die kritische Betrachtung datenbasierter Praktiken sowie die Beurteilung von datengestützten Argumentati-

onen im öffentlichen politischen Diskurs einschließt (etwa im Sinne von Jablonka, 2017; Jablonka, 2024a, 2024b; Jablonka & Barwell, 2024; Jablonka & Bergsten, 2021).

## Anmerkungen

<sup>1</sup> Emenegger (2018) sowie Ludwig und Thiemann (2020) verwenden „Datenkompetenz“ synonym mit *data literacy*. Schüller und Busch (2019) hingegen fassen in ihrer systematischen Zusammenstellung Datenkompetenz[en] als die in *data literacy* enthaltenden Fähigkeiten/Fertigkeiten oder *skills*.

<sup>2</sup> Dieser Datenbegriff ist nur anwendbar auf strukturierte Daten (im Gegensatz zu unstrukturierten Daten) ohne statistische Fehler/Unsicherheiten.

<sup>3</sup> Ridsdale et al. (2015) verstehen *data literacy* als eine Reihe von Kompetenzen und Tätigkeiten, allen voran jedoch „the ability to collect, manage, evaluate, and apply data, in a critical manner“ (S. 3).

<sup>4</sup> Da wir uns für die epistemischen Konnotationen und weniger für die verschiedenen Datenpraktiken interessieren, stützt sich die begriffsgeschichtliche Skizze nur am Rande auf wissenschaftssoziologische Beiträge.

<sup>5</sup> So kriechen bspw. in Aufgaben Würmer, Schlangen oder Löwen mit kontinuierlicher Geschwindigkeit Türme oder Brunnen hinauf und hinunter, wobei die auszurechnende Kriechdauer die Lebenserwartung der Tierarten erheblich übersteigt (vgl. Gellert & Jablonka, 2023).

<sup>6</sup> Die Hauptrepräsentationsform bestand in umfangreichen Tabellen mit variierenden Absolutwerten und Prozentangaben zu verschiedenen Kategorien und Merkmalen, oft aufgeschlüsselt nach Regionen und Zeiträumen, immer mit begleitenden Texten im Argumentationszusammenhang.

<sup>7</sup> Eine Verschiebung des Fokus weg von der beschreibenden Dokumentation durch Datentabellen zur Suche nach Gesetzmäßigkeiten sieht Desrosières (2005) für die Sozialwissenschaften zuerst in England und Frankreich.

<sup>8</sup> Rosenberg bezieht sich auf die lateinische Pluralform in englischsprachigen Texten. Aber auch in Alexander von Humboldts Schriften mit geophysikalischen Themen (Humboldt, 1799) wird allgemein auf "Data" verwiesen, meist ist jedoch von „Werthen“ die Rede.

<sup>9</sup> Berüchtigt ist die bevölkerungspolitische Interpretation des Eugenikers Francis Galton, der die theoretisch angenommene Verteilung (un)erwünschter Fähigkeiten um den Mittelwert kausaldeterministisch als vererbungsbedingt postulierte, während er empirisch festgestellte Abweichungen zufallsbedingten Schwankungen zuschrieb.

<sup>10</sup> Der Informationsbegriff bezieht sich dabei bekanntlich auf die statistischen Eigenschaften des gesamten Ensembles möglicher Nachrichten, die eine bestimmte Art von Quelle produzieren kann, als Maß für die Entscheidungsfreiheit der Quelle bei der Auswahl einer Nachricht.

<sup>11</sup> Damit verlagert sich freilich die Frage nach der Bedeutung für Erkenntnis- und Argumentationsprozesse auf den Informationsbegriff.

<sup>12</sup> Ganz im Sinne einer *data literacy* wäre zu überlegen, wie das Grundprinzip der statistischen Analysen der Textdaten des DWDS zugänglich gemacht werden könnte, die mit „Trankit“, einem „a multi- lingual Transformer-based NLP Toolkit“, erstellt werden (van Nguyen et al., 2021). Für das Verständnis des benutzten Maßes der typischen Kookkurrenzpartner mit

„logDice“ oder ähnlicher Assoziationsmaße (Rychlý, 2008) bietet die Schulmathematik der Sekundarstufe tatsächlich eine Grundlage. Im Übrigen ist der Vergleich solcher Maße ein interessantes mathematisches Modellierungsbeispiel.

<sup>13</sup> Bezogen auf Daten als Grundlage von Evidenz verweist Rosenberg (2014) auf die Etymologie von Evidenz (von Lat. „videre“, sehen können, erkennen, betrachten) und stellt kontrastierend fest, dass „Evidenz“ eine epistemologische, „Daten“ eine argumentativ-rhetorische und „Fakten“ eine ontologische Konnotation haben.

<sup>14</sup> Zum Beispiel des Potsdamer Instituts für Klimaforschung: [https://kfo.pik-potsdam.de//index\\_de.html?language\\_id=de](https://kfo.pik-potsdam.de//index_de.html?language_id=de)

## Literatur

- Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16(1), 3–9.
- Arndt, H. W. (1971). *Methodo scientifica pertractatum, Mos geometrica und Kalkülbegriff in der philosophischen Theoriebildung des 17. und 18. Jahrhunderts*. Walter de Gruyter.
- Aw, A. J. (2014). The mathematics of messages. In S. Parc (Hrsg.), *50 visions of mathematics* (pp. 6–9). Oxford University Press.
- Biehler, R. & Weber, W. (1995). Entdeckungsreisen im Datenland. *Computer und Unterricht*, 17, 4–9.
- Bundesregierung. (2021). *Datenstrategie der Bundesregierung. Eine Innovationsstrategie für gesellschaftlichen Fortschritt und nachhaltiges Wachstum*. Kabinettfassung, 27. Januar 2021.
- Burrill, G. & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill & C. Reading (Hrsg.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education*. New ICMI Study Series, vol 14. Springer. [https://doi.org/10.1007/978-94-007-1131-0\\_10](https://doi.org/10.1007/978-94-007-1131-0_10)
- Christy, J. R. (2016). U. S. House Committee on Science, Space and Technology: Testimony of John R. Christy. <https://www.commerce.senate.gov/services/files/FCBF4CB6-3128-4FDC-B524-7F2AD4944C1D>
- Desai, J., Watson, D., Vincent, W., Mariarosaria, T. & Luciano, F. (2022). The epistemological foundations of data science: a critical review. *Synthese*, 200 (469), 1–27. <https://doi.org/10.1007/s11229-022-03933-2>
- Desrosières, A. (2005). *Die Politik der großen Zahlen: eine Geschichte der statistischen Denkweise*. Übers. Manfred Stern. Springer Web.
- DWDS (2024). „Daten“, bereitgestellt durch das Digitale Wörterbuch der deutschen Sprache <https://www.dwds.de/wb/Daten>
- Emmenegger, S. (2023). Datenkompetenz. Möglichkeiten und Grenzen von Agenda-Settings zu Technikfolgen im Horizont von Bildung. In M. F. Buck & M. Zulaica y Mugica (Hrsg.), *Digitalisierte Lebenswelten: Bildungstheoretische Reflexionen* (S. 231–250). Springer. [https://doi.org/10.1007/978-3-662-66123-9\\_12](https://doi.org/10.1007/978-3-662-66123-9_12)
- Engel, J. (1999). Computer und Erziehung zur Datenkompetenz. In G. Kadunz, G. Ossimitz, W. Peschek, E. Schneider & B. Winkelmann (Hrsg.), *Mathematische Bildung und neue Technologien: Vorträge beim 8. Internationalen Symposium zur Didaktik der Mathematik Universität Klagenfurt, 28.9. –*

- 2.10.1998 (S. 77–84). Vieweg+Teubner Verlag. [https://doi.org/10.1007/978-3-322-90149-1\\_8](https://doi.org/10.1007/978-3-322-90149-1_8)
- Engel, J. (2017). Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal*, 16(1), 44–49.
- Engel, J., Biehler, R., Frischemeier, D., Podworny, S., Schiller, A. & Martignon, L. (2019). Zivilstatistik: Konzept einer neuen Perspektive auf *Data literacy* und Statistical Literacy. *AStA Wirtschafts- und Sozialstatistisches Archiv*, 13(3), 213–244. <https://doi.org/10.1007/s11943-019-00260-w>
- Euclid (1625 / 1896). *Euclidis opera omnia*. Vol. VI Euclidis Data. Zweisprachige Ausgabe übersetzt von J. L. Heiberg & H. Menge. Teubner.
- Feistner, E. (2018). Relativierte Referentialität: Überlegungen zu einer Kulturgeschichte der Interaktion von Erzählen und Rechnen. In E. Feistner (Hrsg.), *Erzählen und Rechnen. Mediävistische Beiträge zur Interaktion zweier ungleicher Kultertechniken* (S. 7–40). Verlag des Bibliotheks- und Informationssystems der Carl von Ossietzky Universität.
- Gal, I. (2002). Adults' Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review / Revue Internationale de Statistique*, 70(1), 1–25. <https://doi.org/10.2307/1403713>
- Gal, I., Nicholson, J. & Ridgway, J. (2022a). A Conceptual Framework for Civic Statistics and Its Educational Applications. In J. Ridgway (Hrsg.), *Statistics for Empowerment and Social Engagement: Teaching Civic Statistics to Develop Informed Citizens* (S. 37–66). Springer International Publishing. [https://doi.org/10.1007/978-3-031-20748-8\\_3](https://doi.org/10.1007/978-3-031-20748-8_3)
- Gal, I., Ridgway, J., Nicholson, J. & Engel, J. (2022b). Implementing Civic Statistics: An Agenda for Action. In J. Ridgway (Hrsg.), *Statistics for Empowerment and Social Engagement: Teaching Civic Statistics to Develop Informed Citizens* (S. 67–96). Springer International Publishing. [https://doi.org/10.1007/978-3-031-20748-8\\_4](https://doi.org/10.1007/978-3-031-20748-8_4)
- Gafny, R. & Ben-Zvi, D. (2024). Reimagining data education: Bridging between classical statistics and data science. In S. Podworny, D. Frischemeier, M. Dvir & D. Ben-Zvi (Hrsgs.) *Minerva School 2022: Reasoning with data models and modeling in the big data era. Universität Paderborn* (S. 69–80). <https://doi.org/10.17619/UNIPB/1-1815>
- Gebre, E. H. (2018). Young Adults' Understanding and Use of Data: Insights for Fostering Secondary School Students' *Data literacy*. *Canadian Journal of Science, Mathematics and Technology Education*, 18(4), 330–341. <https://doi.org/10.1007/s42330-018-0034-z>
- Gebre, E. (2022). Conceptions and perspectives of *data literacy* in secondary education. *British Journal of Educational Technology*, 53. <https://doi.org/10.1111/bjet.13246>
- Gellert, U. & Jablonka, E. (2023). Narrative Räume für den Mathematikunterricht? Eine Spurensuche. In N. Ferrin & I. Kellermann (Hrsg.), *Narrative Räume für das Denken in Möglichkeiten. Perspektivität – Fiktionalität – Kreativität* (S. 55–68).
- Gould, R. (2017). *Data literacy* is statistical literacy. *Statistics Education Research Journal*, 16(1), 22–25. <https://doi.org/10.52041/serj.v16i1.209>
- Gould, R. (2024). Traditional statistical models in a sea of data: teaching introductory data science. In S. Podworny, D. Frischemeier, M. Dvir & D. Ben-Zvi (Hrsg.) *Minerva School 2022: Reasoning with data models and modeling in the big*
- data era. Universität Paderborn (S.81–89). <https://doi.org/10.17619/UNIPB/1-1815>
- Grillenberger, A. & Romeike, R. (2018). Developing a theoretically founded data literacy competency model. *Proceedings of the 13th Workshop in Primary and Secondary Computing Education*, 1–10.
- Grillenberger, A. (2019). Von Datenmanagement zu data literacy: Informatikdidaktische Aufarbeitung des Gegenstandsbereichs Daten für den allgemeinbildenden Schulunterricht. <https://doi.org/10.17169/refubium-1932>
- Hacking, I. (1982). Biopower and the avalanche of printed numbers. *Humanities in Society*, 5(3-4), 279–295.
- Hacking, I. (1987). Prussian numbers 1860–1882. In L. Krüger, L. Daston & L. Heidelberger (Hrsg.), *The probabilistic revolution vol. I*. (S. 377–394). MIT Press.
- Hancock, C., Kaput, J. J. & Goldsmith, L. T. (1992). Authentic inquiry with data: critical barriers to classroom implementation. *Educational Psychologist*, 27(3), 337–364.
- Heinemann, B., Opel, S., Budde, L., Schulte, C., Frischemeier, D., Biehler, R., Podworny, S. & Wassong, T. (2018). Drafting a Data Science Curriculum for Secondary Schools. *Proceedings of the 18th Koli Calling International Conference on Computing Education Research*, 1–5. <https://doi.org/10.1145/3279720.3279737>
- Heintz, B. (2007). Zahlen, Wissen, Objektivität: Wissenschaftssoziologische Perspektiven. In A. Mennicken & H. Vollmer (Hrsg.), *Zahlenwerk: Kalkulation, Organisation und Gesellschaft* (S. 65–85). VS Verlag für Sozialwissenschaften.
- Hrycej, T., Bermeitinger, B., Cetto, M. & Handschuh, S. (2023). *Mathematical Foundations of Data Science*. Springer Nature.
- Humboldt, A. (1799). Versuche über die chemische Zerlegung des Luftkreises und über einige andere Gegenstände der Naturlehre. Vieweg.
- Jablonka, E. (2017). Mathematisation in environments of Big Data – “implicit mathematics” revisited. In B. di Paola & U. Gellert (Hrsg.), *Proceedings from CIEAEM 69. Quaderni di Ricerca in Didattica (Mathematics)*, 27(Supplemento n. 2), 43–51.
- Jablonka, E. (2024a). Die Auseinandersetzung mit Mathematik im öffentlichen Diskurs als Teil eines politisch bildenden Mathematikunterrichts? In T. Hamann, M. Helmerich, D. Kollosche, K. Lengnink & S. Pohlkamp (Hrsg.), *Mathematische Bildung neu denken: Andreas Vohns erinnern und Weiterdenken* (S. 83–98). WTM-Verlag.
- Jablonka, E. (2024b). Reading Koopman's (2019) “How we became our data” as an invitation to resist the formatting of the “informational person” with the support of mathematics education. *Journal of the American Association for the Advancement of Curriculum Studies*, 16(1), 91–106. <https://ojs.library.ubc.ca/index.php/jaaacs/article/view/198586/192954>
- Jablonka, E. & Barwell R. (2024). Critical thinking in STEM education. In L. D. English & T. Lehmann (Hrsg.), *Ways of thinking in STEM-based problem solving: Teaching and learning in a new era* (S. 192–204). Routledge. <https://doi.org/10.4324/9781003404989>
- Jablonka, E. & Bergsten, C. (2021). Numbers don't speak for themselves: strategies of using numbers in public policy discourse. *Educational Studies in Mathematics*, 108(3), 579–596. <https://doi.org/10.1007/s10649-021-10059-8>

- KMK (2023). Bildungsstandards MATHEMATIK (2022) Primarstufe und Sekundarstufe I. Beitrag zur Implementation. Sekretariat der Kultusministerkonferenz.
- KomSZ-Kommission Zukunft Statistik (2024). Bericht und Empfehlungen der Kommission Zukunft Statistik. Bericht 15.01.2024.
- Königl. Preuß. Statist. Bureau (1860/61). *Zeitschrift des Königlich Preussischen Statistischen Bureaus*. Verl. d. Königl. Statist. Bureaus.
- Lee, V. R., Drake, J., Cain, R. & Thayne, J. (2021). Remembering what produced the data: Individual and social reconstruction in the context of a Quantified Self elementary data and statistics unit. *Cognition and Instruction*, 39(4), 368–408. <https://doi.org/10.1080/07370008.2021.1936529>
- Ludwig, T. & Thiemann, H. (2020). Datenkompetenz – *Data literacy*. *Informatik Spektrum*, 43(6), 436–439. <https://doi.org/10.1007/s00287-020-01320-0>
- Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S. & Wuetherick, B. (2015). *Strategies and Best Practices for Data literacy Education: Knowledge Synthesis Report*. <https://doi.org/10.13140/RG.2.1.1922.5044>
- Röhner, G., Brinda, T., Denke, V., Hellwig, L., Heußer, T., Pasternak, A. & Seiffert, M. (2016). Bildungsstandards Informatik für die Sekundarstufe II. Beilage zu LOG IN, 36(183/184)
- Rosenberg, D. (2014). In R. Reichert (Hg.). *Big Data. Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie* (S. 133–156). transcript Verlag.
- Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, 33. <https://doi.org/10.1177/0165551506070706>
- Rychlý, P. (2008, December). A Lexicographer-Friendly Association Score. In P. Sojka, A. Horák (Hrsg.), *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN RASLAN* (pp. 6-9). Masaryk University, Brno.
- Schüller K., Busch P. & Hindinger C. (2019). Future Skills: Ein Framework für *Data Literacy*. *Hochschulforum Digitalisierung* (Arbeitspapier 47/2019)
- Schüller, K., Koch, H. & Rampelt, F. (2021). *Data-Literacy-Charta*. Berlin: Stifterverband für die Deutsche Wissenschaft, Stifterverband. <https://www.stifterverband.org/charter-data-literacy>
- SenBJF-Senatsverwaltung für Bildung, Jugend und Familie Berlin; Ministerium für Bildung, Jugend und Sport des Landes Brandenburg (2023). Rahmenlehrplan Teil C Mathematik Jahrgangsstufen 1 – 10.
- Shannon, C. E. & Weaver, W. (1949/1964). *The mathematical theory of communication. Tenth printing*. The University of Illinois Press.
- Sidoli, N. (2018) The concept of given in Greek mathematics. *Arch. Hist. Exact Sci.* 72, 353–402.
- Swijtink, Z. G. (1987). The objectification of observation: measurement and statistical methods in the nineteenth century. In L. Krüger, L. Daston & L. Heidelberger (Hrsg.), *The probabilistic revolution vol. I*. (S. 261–286). MIT Press.
- Van Nguyen, M., Lai, V. D., Veyseh, A. P. B. & Nguyen, T. H. (2021). Trankit: A light-weight transformer-based toolkit for multilingual natural language processing. arXiv preprint arXiv:2101.03289.
- Wild, C. J. (2017). Statistical literacy as the earth moves. *Statistics Education Research Journal*, 16(1), 31–37.
- Wolff, A., Gooch, D., Caverio, M., Jose J., Rashid, U. & Kortuem, G. (2017). Creating an understanding of *data literacy* for a data-driven society. *The Journal of Community Informatics*, 12(3), 9–26. <https://doi.org/10.15353/joci.v12i3.3275>
- Zhang, K., Liu, S. & Xiong, M. (2022). *Changes from classical statistics to modern statistics and data science*. arXiv preprint arXiv:2211.03756. <https://doi.org/10.48550/arXiv.2211.03756>

## Anschrift der Verfasserinnen

Eva Jablonka  
Freie Universität Berlin  
Fachbereich Erziehungswissenschaft und Psychologie  
Habelschwerdter Allee 45  
14195 Berlin  
[eva.jablonka@fu-berlin.de](mailto:eva.jablonka@fu-berlin.de)

Katharina Jablonka-Cohen  
Freie Universität Berlin  
Fachbereich Erziehungswissenschaft und Psychologie  
Habelschwerdter Allee 45  
14195 Berlin  
[k.jablonka@fu-berlin.de](mailto:k.jablonka@fu-berlin.de)