

# Datenbewusstsein: Aufmerksamkeit für die eigenen Daten

Lukas Höper,<sup>1</sup> Carsten Schulte<sup>1</sup>

**Abstract:** Die aktuellen Diskurse über künstliche Intelligenz, maschinelles Lernen, Big Data und Data Science zeigen die zunehmende Bedeutung von Daten in der digitalen Gesellschaft. Informatiksysteme, die Technologien der künstlichen Intelligenz oder spezieller des maschinellen Lernens verwenden, sammeln und verarbeiten in der Regel ständig Daten ihrer Nutzer\*innen. Das fehlende Bewusstsein für die Sammlung und Verarbeitung der eigenen, in der Regel personenbezogenen, Daten bedeutet ein fehlendes Verständnis von diesen genutzten Informatiksystemen. Ein bewusster und kompetenter Umgang mit datengetriebenen Informatiksystemen wird somit Grundlage jeder informationellen Selbstbestimmung und ist damit Voraussetzung für eine selbstbestimmte Teilhabe an der digitalen Gesellschaft. Unser Beitrag stellt zu diesem Zwecke das Konzept Datenbewusstsein vor, womit wir auf neuartige Weise verschiedene Inhalte und Kompetenzen zu einem Gesamtkonzept verbinden.

**Keywords:** Data Science; Maschinelles Lernen; Allgemeinbildung; Datenbewusstsein; Datengetriebene Informatiksysteme

## 1 Einleitung

In diesem Artikel stellen wir das neue Konzept *Datenbewusstsein* vor. Damit reagieren wir auf die gestiegene Bedeutung datengetriebener Anwendungen und Ausdifferenzierungen der Informatik, wie sie etwa im Bereich Data Science, Big Data und datengetriebenem Lernen sowie in verschiedenen Anwendungsfeldern, wie z.B. Datenjournalismus, Open Data oder Data Society anklängen. Beim Nutzen der meisten Informatiksysteme werden immer häufiger und in immer größerem Umfang Daten gesammelt. Ständig und überall hinterlassen wir bewusst oder unbewusst Daten. Heutzutage finden sich im Alltag nahezu überall Informatiksysteme, die Daten der Nutzer\*innen sammeln und verarbeiten. Mit dem Konzept *Datenbewusstsein* schlagen wir einen Ansatz vor, wie Lernende für die Sammlung und Verarbeitung der eigenen, in der Regel personenbezogenen, Daten aufmerksam gemacht werden können, um bei ihnen ein Bewusstsein für die Prozesse der Sammlung und Verarbeitung ihrer Daten zu erzeugen. Ein solches Datenbewusstsein ist Voraussetzung für einen bewussten und kompetenten Umgang mit datengetriebenen Informatiksystemen und somit einer selbstbestimmten Teilhabe an der digitalen Gesellschaft. Im Sinne der informationellen Selbstbestimmung hat Datenbewusstsein demnach das Potenzial, ein allgemeinbildendes Ziel des Informatikunterrichts zu sein. Die hinter Datenbewusstsein stehenden Ideen sind einzeln betrachtet nicht gänzlich unbekannt oder neu. Innovativ

---

<sup>1</sup> Universität Paderborn, Didaktik der Informatik, Fürstenallee 11, D-33102 Paderborn, {lukas.hoepfer,carsten.schulte}@upb.de

hingegen ist ihre Kombination. Das Konzept greift verschiedene Ansätze auf, die wir im folgenden Abschnitt ‚Stand der Forschung‘ einführen. Anschließend werden datengetriebene Informatiksysteme und die Interaktion der Nutzer\*innen mit diesen Systemen beleuchtet. Schließlich wird in diesem Kontext das Reflektieren sowohl auf einer individuellen als auch auf einer gesellschaftlichen Ebene beschrieben, um daraus das Konzept und einen konkreten Rahmen für die Umsetzung in der informatischen Allgemeinbildung, das heißt insbesondere im schulischen Informatikunterricht, zu entwickeln.

## 2 Stand der Forschung und verwandte Ansätze

Bei der Nutzung nahezu aller datengetriebenen Informatiksysteme werden auch Daten von Nutzer\*innen erhoben. Anbietende solcher datengetriebener Informatiksysteme tracken ständig ihre Nutzer\*innen, um ein Profiling durchzuführen und um schließlich wirtschaftliche Gewinne zu erzielen [Zu18]. Dabei werden Erfahrungen der Nutzer\*innen als kostenloser Rohstoff für kommerzielle Zwecke genutzt, wie etwa für Vorhersagen des zukünftigen Verhaltens der Nutzer\*innen. Zuvor wurden Daten gesammelt, um Dienste anbieten zu können, nun aber werden die Dienste angeboten, um gerade diesen *Verhaltensüberschuss* [Zu18, S. 22] abgreifen zu können. Nach [Zu18] entsteht dadurch insgesamt eine neue Form des Kapitalismus: Der *Überwachungskapitalismus*. Die sozialen und gesellschaftlichen Implikationen solcher Systeme sind immens und betreffen nicht nur das Wirtschaftssystem [Ra19]. Insgesamt lässt sich die Entwicklung als eine Datafizierung der Gesellschaft beschreiben [Ma17]. Brisant wird der angestoßene Wandel insbesondere auch dadurch, dass die Systeme nur schwer zu durchschauen, gleichsam undurchsichtig sind [Bu16], die Umwälzungen viele Bereiche des Lebens und der Gesellschaft betreffen, und zu ihrer Bewältigung daher auch neue gesellschaftliche Regulierungen gefunden werden müssen, die nicht allein durch die Technik oder die IT-Industrie festgelegt werden [Ma17; SS18]. Aus Sicht der Informatik als Disziplin bedeutet das möglicherweise nicht nur eine neue Herangehensweise, sondern die Notwendigkeit für eine eigene Disziplin mit eigenen Fragen und methodischen Zugängen, etwa um das Verhalten der ‚intelligenten Maschinen‘ auf verschiedenen Ebenen zu erforschen [Ra19].

**Bildungsansätze als Reaktionen** Als Reaktion auf die genannten Entwicklungen sind verschiedene Bildungsansätze entwickelt worden, etwa Data Literacy, Statistical Literacy oder AI Literacy. Mit Data Literacy wird nach [Ri15] die Fähigkeit beschrieben, „Daten auf kritische Art und Weise zu sammeln, zu managen, zu bewerten und anzuwenden“ [Ri15]<sup>2</sup>. Statistical Literacy fokussiert sich auf statistische Fragestellungen [Ga02]. AI Literacy bezieht sich wiederum auf eine breite Kompetenzsammlung, unter anderem zum Evaluieren von Technologien der künstlichen Intelligenz und Nutzen dieser als Tool [LM20]. Stärker auf den allgemeinbildenden Informatikunterricht bezogen ist das Konzept Data Agency nach

---

<sup>2</sup> Übersetzung von: <https://hochschulforumdigitalisierung.de/de/dossiers/data-literacy>, zuletzt aufgerufen am 09.07.2021

Tedre et al., welches als Weiterentwicklung von Data Literacy beschrieben wird [Te20]. Sie beschreiben Data Agency als: „people’s volition, skills, attitudes, and capacity for informed actions that make a difference in their digital world“ [Te20]. Damit zielen die Autor\*innen auf eine informierte Handlungsfähigkeit im Sinne eines selbstbestimmten Handelns in der digitalen Welt ab. In den verschiedenen Ansätzen werden unterschiedliche Aspekte des Umgangs mit Daten betrachtet, doch ein Fokus auf die Sammlung und Verarbeitung der eigenen Daten während dem Nutzen von datengetriebenen Informatiksystemen fehlt allerdings bisher.

**Datensammlung im Rahmen der informationellen Selbstbestimmung** Aus dem Datenschutzrecht entstammt das Konzept der informationellen Selbstbestimmung, die dem Menschen ermöglichen soll, frei über das Hinterlassen bzw. Sammeln-Lassen ihrer personenbezogenen Daten und deren Verarbeitung bestimmen zu können.<sup>3</sup> Es stellt sich die Frage, wie diese aussehen kann, wenn den Nutzer\*innen verborgen bleibt, was mit ihren Daten geschieht, sofern sie überhaupt wissen, dass ihre Daten gesammelt werden. Den Nutzer\*innen müsste also bewusst sein, dass Daten gesammelt und wie und wozu sie verarbeitet werden. Zwischen dem bewussten Hinterlassen von Daten und dem womöglich nicht offensichtlichen und damit der Nutzerin und dem Nutzer nicht bewussten Sammeln der Daten seitens datengetriebener Informatiksysteme sollte unterschieden werden; etwa nach der Kategorisierung der OECD, die folgende Arten von Daten unterscheidet: (a) *provided data* als Daten, die bewusst und aktiv durch die Nutzer\*innen erstellt werden; (b) *observed data* als Beobachtungs- und Metadaten, deren Erhebung dem Einzelnen nicht bewusst sein muss; (c) *derived data* als Daten, die aus anderen bestehenden Daten generiert wurden; (d) *inferred data* als Daten, die durch wahrscheinlichkeitsbasierte analytische Prozesse generiert werden, wie etwa Vorhersagewerte für das Verhalten des Nutzenden [OE14] (ähnlich: [PS19]). Im Sinne der informationellen Selbstbestimmung reicht es also nicht aus, nur zu wissen, *dass* Daten gesammelt werden. Für einen kompetenten Umgang mit datengetriebenen Informatiksystemen muss den Nutzer\*innen vielmehr auch bewusst gemacht werden, wie und wozu ihre Daten verwendet und verarbeitet werden. Eine Untersuchung von [PS19] zeigt, dass sich Schüler\*innen zwar bewusst sind, dass sie Systeme nutzen, welche etwa Methoden des maschinellen Lernens verwenden; es fehlt jedoch das Verständnis, wo, wie und wozu Daten über sie gesammelt und verarbeitet werden, jedoch fehlt (vgl. auch [Te20]). Um ein solches Verstehen zu entwickeln, ist jedoch auch eine Verschiebung bei der Thematisierung der Systeme mit Fokus auf Algorithmen und Datenstrukturen hin zur Sammlung, Verarbeitung und Organisation von Daten nötig [SFN18]. Dieser Schritt sollte dazu führen, dass die Schüler\*innen datengetriebene Informatiksysteme verstehen können.

<sup>3</sup> vgl. <https://www.bfdi.bund.de/SharedDocs/Downloads/DE/Flyer/Datenschutz-ist.html>, zuletzt aufgerufen am 09.07.2021

### 3 Das Konzept Datenbewusstsein

In diesem Abschnitt stellen wir schrittweise einzelne Facetten des Konzepts *Datenbewusstsein* vor, um dieses Konzept anschließend samt einem konkreten Rahmen für die Umsetzung in der informatischen Allgemeinbildung zu erläutern.

#### **Datengetriebene Informatiksysteme und datengetriebene digitale Artefakte (ddA)**

Viele Informatiksysteme setzen Verfahren des maschinellen Lernens oder allgemeiner der künstlichen Intelligenz ein, um die gesammelten Daten (Big Data) zu verarbeiten. Ein solches Informatiksystem nennen wir datengetrieben. Die Nutzer\*innen nutzen in einer konkreten Situation nur einen Teil des Informatiksystems, den wir als digitales Artefakt bezeichnen, wie etwa der News Feed auf einer Social Media Plattform. Da die Nutzer\*innen in der Regel ein konkretes digitales Artefakt nutzen und nicht das gesamte datengetriebene Informatiksystem, werden wir uns im Folgenden primär auf *datengetriebene digitale Artefakte (ddA)* beziehen, wobei diese immer als Teil eines Informatiksystems zu verstehen sind. Für viele Anbietende eines ddA sind oft vor allem personenbezogene Daten interessant, um etwa Informationen über die Person zu gewinnen und deren zukünftiges Verhalten vorherzusagen. Die Nutzer\*innen können diese beim Nutzen eines ddA bewusst hinterlassen, etwa die E-Mail Adresse beim Erstellen eines Accounts auf einer Plattform. Darüber hinaus werden jedoch während der Nutzung Daten gesammelt, worüber sich die Nutzer\*innen nicht unbedingt bewusst sein müssen, da sie für die mit dem Artefakt durchgeführte Handlung selbst nicht primär relevant sind. Ein gutes Beispiel hierfür sind Suchmaschinen, bei denen nicht nur die von der Nutzerin oder dem Nutzer bewusst eingegebenen Suchbegriffe sondern auch jeder Klick auf ein Suchergebnis abgespeichert wird. Damit ein selbstbestimmter Umgang mit einem ddA auch mit Daten, die während der Nutzung erhoben (observed data) oder durch eine Verarbeitung erzeugt werden (derived und inferred data), möglich ist, muss diesbezüglich eine Aufmerksamkeit für und ein adäquates Verständnis des ddAs im Nutzungskontext gesorgt werden; ein solches fehlt jedoch zumeist [PS19].

#### **Interaktion zwischen Nutzer\*innen und datengetriebenen digitalen Artefakten**

Damit die ddA im Sinne eines Datenbewusstseins verstanden werden können, sollten sie nicht nur rein objektiv und kontextfrei in ihrer technischen Beschaffenheit beschrieben werden, sondern vielmehr im jeweiligen Nutzungs- bzw. Interaktionskontext betrachtet werden. Eine Interaktion kann als eine Kette von Eingaben, Verarbeitungen und Ausgaben verstanden werden [SB18]. Die Nutzerin oder der Nutzer tätigt eine Eingabe, wie etwa das Anklicken oder Bewerten eines Produktes. Die Eingabe kann dann auf zweierlei Weise interpretiert werden: 1) Die Eingabe ist eine Aufforderung an das ddA etwas zu tun und 2) durch die Eingabe werden bewusst oder unbewusst Daten hinterlassen. Sowohl die Aufforderung als auch die Daten werden zur Erzeugung einer Ausgabe verarbeitet. So können etwa Vorhersagen für Interessen der Nutzer\*innen ermittelt und damit personalisierte Ausgaben erzeugt werden. Dem folgt eine weitere Eingabe seitens der Nutzerin oder des Nutzers,

was dem typischen Charakter einer Interaktion entspricht. Die Daten werden also während einer Interaktion zwischen einer Nutzerin oder einem Nutzer mit einem ddA erhoben bzw. entstehen durch und in der Interaktion. Die Lernenden kennen Interaktionen dieser Art bereits aus ihrem Alltag und haben dazu eigene Erfahrungen gesammelt. Sie sollten diese Interaktionen und die möglicherweise stattfindende, implizite Datensammlung durch das ddA verstehen und dafür aufmerksam werden, weshalb der Interaktion in diesem Ansatz eine wichtige Rolle zukommt. Für das Konzept Datenbewusstsein ist nun folgender Aspekt der Interaktion besonders wichtig: Bei diesen Interaktionen entsteht eine Wechselwirkung oder auch gegenseitige Beeinflussung [SB18]. Einerseits können Nutzer\*innen zum Beispiel durch personalisierte Werbung etwa im Sinne des Verleitens zum Anklicken oder Kaufen eines Produktes manipuliert werden. Andererseits wird ihr Verhalten protokolliert und ausgewertet und kann so zu einem angepassten Vorschlag für nachfolgende Interaktionen (u.a. auch durch andere Nutzer\*innen) führen. Das heißt, Interaktionen und die dabei anfallenden Daten können einerseits die Nutzer\*innen, andererseits das System beeinflussen. Diese wechselseitigen Beeinflussungsmöglichkeiten finden sich in vielen Anwendungsfeldern, wie beispielsweise den von [Ra19] genannten, von denen wir nur das folgende benennen werden: Filterblasen, die die Nutzer\*innen in Folge dieser Wechselwirkungen ausprägen und durch diese beeinflusst werden. Die Nutzer\*innen können ebenfalls durch bewusste Handlungen – wie etwa das absichtliche Hinterlassen oder das Verhindern der Sammlung von Daten – versuchen, das ddA zu manipulieren.

**Duales Verständnis datengetriebener digitaler Artefakte** Die Interaktionen zwischen einer Nutzerin oder einem Nutzer und einem ddA samt der diesbezüglich bestehenden Wechselwirkung sollte aus der Perspektive einer individuellen und einer gesellschaftlichen Betrachtungsebene thematisiert werden. Zum einen sollten aus der technischen bzw. architekturbezogenen Perspektive die Mechanismen und Verfahren der Verarbeitung der Daten zumindest ansatzweise begreifbar werden – diese nennen wir analytische Perspektive der Architektur. Das Verstehen der Architektur ist eine Voraussetzung, um zum anderen über die Bedeutsamkeit und die Auswirkungen dieser Prozesse zu sprechen, was zum Verstehen und insbesondere Reflektieren dieser Interaktionen und Wechselwirkungen ebenfalls nötig ist – diese nennen wir interpretierende Perspektive der Relevanz (vgl. [SB18]). Damit die Lernenden ein ddA gänzlich begreifen können, müssen beide Perspektiven eingenommen werden. Die Relevanz eines ddA lässt sich ohne die Architektur nicht begründen und die Architektur kann wiederum ohne die Relevanz nicht verstanden werden, da etwa Aspekte wie die Fairness eines datengetriebenen Informatiksystems allein aus der Architektur schwer zu verstehen sind. Die Architektur sollte zum Erreichen eines Verständnisses von ddA also nicht losgelöst vom Bereich *Informatik, Mensch und Gesellschaft* betrachtet werden.

**Datenmodelle: Digitale Doppelgänger** Es stellt sich nun die Frage, was es bedeutet, dass Daten verarbeitet werden. Es werden Datenmodelle erzeugt, die sowohl die Daten der Nutzerin bzw. des Nutzers in dem Interaktionskontext als auch die Daten aller anderen

Nutzer\*innen samt anderer Datenmodelle einbeziehen. Ein Beispiel dafür ist das Konzept des digitalen Doppelgängers [BK15]. Ähnliche Konzepte sind etwa digitale Profile, digitale Identitäten, digitale Double oder Datenschatten (z.B. [Ma17], [Be15]). Ein digitaler Doppelgänger ist nie vollständig fertig und kann sich daher mit jeder neuen Dateneingabe ändern [BK15]. Er spiegelt dabei die Nutzerin bzw. den Nutzer wider, wobei eine Abhängigkeit zu ihr bzw. zu ihm entsteht [BK15]. Insgesamt kann der digitale Doppelgänger anhand verschiedener Merkmale als digitale Charakterisierung einer Person gesehen werden. Diese Merkmale, das ist wichtig zu begreifen, bestehen nur zu einem eher kleineren Teil aus den provided und observed data; sie bestehen vor allem auch aus derived und inferred data. Um das deutlich zu machen und ein entsprechendes Datenbewusstsein zu entwickeln, können diese derived und inferred data generierenden Extraktions- und Veredelungsprozesse (vgl. [Zu18]) im Unterricht von Lernenden explorierend selbst durchgeführt werden. Dazu reichen schon recht einfache Veredelungsprozesse, wie Filtern und Visualisieren, die als Beispielvorgang im Unterrichtskontext anschaulich die Mächtigkeit der Datenverarbeitung, aber auch deren Ungenauigkeiten, aufdecken können (vgl. Abs. 5). Interessant ist zudem, dass diese Datenmodelle, die als Ergebnis der Verarbeitung von Daten entstehen, in einer konkreten Interaktion zwischen einer Nutzerin oder einem Nutzer und dem ddA erstellt und so etwa zu Vorhersagen des Nutzerverhaltens herangezogen werden.

**Reflexion auf individueller und gesellschaftlicher Ebene** Im Rahmen von Datenbewusstsein sehen wir eine Reflexion auf individueller Ebene als wichtig an, um die eigene Rolle in dieser Interaktion und die bestehende Wechselwirkung zu identifizieren. So sollten insbesondere die eigene Rolle, die eigenen Handlungsmuster und die gegenseitigen Auswirkungen reflektiert und bewertet werden, was eine Voraussetzung für ein selbstbestimmtes Handeln ist. Dabei kann und sollte die Möglichkeit sowohl des Beeinflussens des ddA als auch des Beeinflusst-Werdens angesprochen werden. Außerdem ist es im Rahmen so einer Interaktion bedeutsam, die Interaktion und das Interaktionssystem, bestehend aus den beiden Akteuren – eine Nutzerin bzw. ein Nutzer und das ddA – und der Interaktion zwischen diesen, zu erkennen und zu reflektieren (vgl. [SB18]); denn nur so können die eigene Rolle sowie Einfluss- und Manipulationsmöglichkeiten aufgedeckt werden. Neben der individuellen Ebene ist es ein weiterer Schritt, die Größe der hinter den ddA stehenden Informatiksystemen – etwa bei Suchmaschinen, sozialen Netzwerken oder im Online-Handel – in den Blick zu nehmen und zu überlegen, inwiefern deren massenhaftes Nutzen und die damit verknüpften riesigen, derived und inferred data über die unmittelbaren Daseins-Zwecke der Infrastrukturen hinaus gesellschaftliche Relevanz entfalten. Diese Aspekte werden bei den eingangs genannten Autor\*innen zum Teil sehr dramatisch beschrieben (vgl. etwa [Be15; Ma17; Ra19; Te20; Zu18]). So sollten die Auswirkungen dieser Systeme auf die Gesellschaft thematisiert und mit dem Verständnis der Architektur verknüpft werden. Nur mit einem gewissen Verständnis der technischen Hintergründe lassen sich die Grenzen und Chancen des Einsatzes eines ddAs beurteilen und begründen. Durch das Reflektieren auf einer individuellen und gesellschaftlichen Ebene in diesem Zusammenhang sollen die Lernenden dazu befähigt werden, datengetriebene Informatiksysteme adäquat bewerten zu können

und somit eigene Entscheidungen im Rahmen einer informationellen Selbstbestimmung zu treffen.

## 4 Datenbewusstsein als Schlussfolgerung

Zuvor wurden verschiedene Facetten des Konzepts Datenbewusstsein vorgestellt, die im folgenden Modell in Abb. 1 zusammengeführt und veranschaulicht werden. Dieser Abschnitt widmet sich zu diesem Zweck der Formulierung wesentlicher Ziele für die Förderung von Datenbewusstsein im Informatikunterricht.

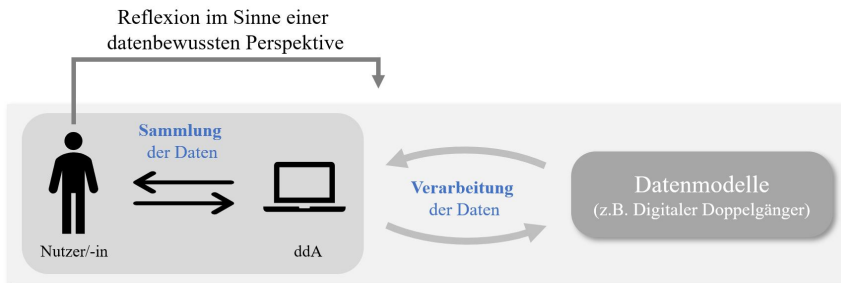


Abb. 1: Modell zur Beschreibung der verschiedenen Facetten von Datenbewusstsein

Für einen bewussten und kompetenten Umgang mit einem ddA sind die Interaktionen zwischen einem Menschen und einem ddA relevant. Diese beiden Akteure lassen sich zusammen mit der Interaktion als Interaktionssystem beschreiben, in dem Daten gesammelt werden. Dies sollte den Nutzer\*innen bewusst sein. Die Zwecke der Datenverarbeitung können im Wesentlichen in zwei Bereiche unterteilt werden: 1) Daten werden zum Betreiben von Funktionen des ddA verarbeitet (primär) oder 2) um etwa Weiterentwicklungen des ddA zu untersuchen oder weitergehende (z.B. kommerzielle) Zwecke zu verfolgen (sekundär). Dabei werden in der Regel Datenmodelle erstellt (etwa ein digitaler Doppelgänger), die zur Verarbeitung neuer erhobener Daten genutzt werden können. Den Nutzer\*innen muss für einen bewussten und kompetenten Umgang mit diesen ddA im Rahmen eines solchen Interaktionssystems die Datensammlung und -verarbeitung bewusst sein. Sie sollten grundlegend die Datenarten (explizit oder implizit erhoben bzw. generiert) sowie Prozesse bzw. Arten und Ergebnisse der Datenverarbeitung kennen, insbesondere deren Zusammenführung zu Datenmodellen und digitalen Doppelgängern. Um diese Prozesse der Sammlung und Verarbeitung von Daten in entsprechenden Interaktionskontexten zu verstehen, muss sowohl eine analytische Perspektive auf die Architektur als auch eine interpretierende Perspektive auf die Relevanz eingenommen und miteinander verbunden werden [Ra19; SB18]. Erst so kann ein Verstehen der Architektur des ddAs und dessen Bedeutungen und Auswirkungen (Relevanz) in einem derartigen Interaktionskontext gelingen, reflektiert und bewertet werden, was für einen bewussten und kompetenten Umgang essentiell ist. Außerdem kann durch die Verknüpfung der beiden Perspektiven ein Verständnis für die Wechselwirkungen in

einem solchen Interaktionssystem vermittelt werden. Ein wesentlicher Aspekt ist, dass es hier insbesondere auch um die individuelle Handlungsfähigkeit geht, die durch das Wechselwirkungsverhältnis von Beeinflussen und Beeinflusstwerden in einem Interaktionssystem mit dem ddA unweigerlich stattfindet (vgl. [SB18]). Neben den Risiken der ggf. unerwünschten Manipulation ergeben sich hier auch vielversprechende und neuartige Handlungsmöglichkeiten, die Tedre et al. als *virtuelle Metakognition* bezeichnen: Als Teilkompetenzen für Datenbewusstsein (sie nennen es data agency) schlagen sie vor, den eigenen digitalen Doppelgänger (digitales Selbst) (a) als Erweiterung und auch als Unterschied zum physischen Selbst verstehen; (b) kritisch bewerten, verbessern und rekonstruieren; und (c) inmitten derer anderer differenzieren und verorten zu können [Te20, Tabelle 1].

## 5 Ausblick auf eine mögliche Umsetzung im Informatikunterricht

Nachdem zuvor das Konzept Datenbewusstsein erläutert wurde, wird in diesem Abschnitt zur Veranschaulichung der wesentlichen Ziele zur Förderung von Datenbewusstsein eine mögliche Unterrichtsskizze einer bereits mehrmals in verschiedenen Klassenstufen erprobten Unterrichtseinheit *Exploration von Standortdaten* beschrieben. Die Unterrichtseinheit für die Sekundarstufe I thematisiert Standortdaten im Mobilfunknetz und wieso diese entstehen und notwendig sind (primäre Funktion). In einem zweiten Schritt werden diese Daten durch die Lernenden explorierend untersucht, mit der Aufgabe anhand der Bewegungsdaten einer Person möglichst viel über diese im Sinne eines digitalen Doppelgängers herauszufinden (als mögliche sekundäre Funktion). Die Kernelemente der Unterrichtseinheit sind im Wesentlichen eine Einstiegsphase, eine Datenexploration sowie eine gemeinsame Auswertung der Ergebnisse. Zunächst wird die Funktionsweise des Mobilfunknetzes eingeführt, etwa durch die Klärung der Frage: Wie kommt ein Anruf zwischen zwei Handys zustande? Im Rahmen der Einstiegsphase wird gemeinsam ergründet, dass Standortdaten durch den Mobilfunkanbieter gesammelt werden, um etwa die Verbindung für einen Anruf gezielt an die passende Basisstation schicken zu können, ohne das gesamte Netz mit Anfragen fluten zu müssen. Eine weitere Funktion dieser Datenerhebung ist, Kosten für Mobilfunkverbindungen anhand der Standorte (etwa länderübergreifend oder in einer Home Zone) berechnen zu können. Die Erhebung der Daten ist also notwendig und zunächst einmal unverfänglich. Anschließend explorieren die Lernenden einen veröffentlichten Datensatz mit Standortdaten aus dem deutschen Mobilfunknetz<sup>4</sup> und charakterisieren die ihnen unbekannt Person, indem sie einen Steckbrief über sie erstellen. Hierzu nutzen sie eine selbstentwickelte interaktive Web-Anwendung oder alternativ ein Jupyter Notebook. Damit können durch Filtern auf zeitliche Ausschnitte (etwa: nachts zwischen 2 und 4 Uhr) und Visualisieren der Datenpunkte auf einer Karte Schlussfolgerungen gezogen werden (etwa: wohnt an dieser Adresse). Die von den Lernenden gewonnenen bzw. interpretierten Informationen (implizite Daten) über die Person werden anschließend in der Auswertung diskutiert und zusammengetragen. Dabei sind vor allem unterschiedliche Interpretationen – wie etwa mögliche, unterschiedliche

---

<sup>4</sup> vgl. <https://www.zeit.de/digital/datenschutz/2011-02/vorratsdaten-malte-spitz>, zuletzt aufgerufen am 09.07.2021



Berufe – interessant. Anschließend werden unter anderem individuelle und gesellschaftliche Ebenen reflektiert und das Interaktionssystem bewertet.

## 6 Fazit und Diskussion

Das Konzept Datenbewusstsein zeigt auf, dass und wie eine Aufmerksamkeit für die Sammlung und Verarbeitung von Daten im Kontext der Interaktion zwischen Nutzer\*innen und datengetriebener digitaler Artefakte (ddA) geschaffen werden kann. Für eine Förderung von Datenbewusstsein im Informatikunterricht resultiert als Ziel eine Fähigkeit der Lernenden in konkreten Interaktionen mit einem ddA sich selbst im Rahmen dieses Interaktionssystems wahrzunehmen und zu überlegen, wo, wie und wozu Daten gesammelt und verarbeitet werden. Dazu ist ein Verständnis über die Prozesse im ddA nötig, um beispielsweise technische Gründe erklärt zu können, warum welche Daten von dem ddA zum Betreiben von angebotenen Funktionen gesammelt und verarbeitet werden müssen. Das Konzept Datenbewusstsein zeigt auf, wie diese, zum Teil in den verschiedenen Empfehlungen zu den Bildungsstandards von der GI enthaltenen, Inhalte und Kompetenzen auf neuartige Weise miteinander verknüpft werden können und sollten. Datenbewusstsein ist für verschiedene Jahrgangsstufen relevant und interessant. Bemerkenswert ist die Aufnahme in dem bald erscheinenden Kernlehrplan von Nordrhein-Westfalen, insofern diese aufzeigt, dass das Konzept hilft, den Informatikunterricht in diesem Kontext zu beschreiben und zu planen. Wir meinen, dass das Konzept Datenbewusstsein eine tragende Säule in der Diskussion um informatische Bildung und deren Weiterentwicklung hin zu einem Pflichtfach verankerten Schulunterricht werden sollte. Ähnlich wie Tedre et al. argumentieren, nimmt es nicht nur die gesellschaftliche Seite der Datenrevolution auf, sondern verknüpft die informatische Bildung auch mit dem ‚data-driven turn‘ der Technologie und der Wissenschaft Informatik [Te20]. Datenbewusstsein hebt sich insbesondere von klassischen Ansätzen der Statistical oder Data Literacy durch diesen Einbezug der Technologie und der damit verknüpften Verfahren ab. Zudem ist es individuell sehr bedeutsam, wie das Konzept des digitalen Doppelgängers verdeutlicht. Dazu ist die Fokussierung auf Interaktionsprozesse u.E. unabdingbar. Möglicherweise ist Datenbewusstsein auch ein wichtiger Baustein, um ddAs und die damit verknüpften Verfahren – etwa des maschinellen Lernens – zu verstehen. Register und Ko [RK20] zeigen, dass sich Lernende an Konzepte des maschinellen Lernens besser erinnern und diese nutzen können, wenn sie zuvor eigene persönliche Daten eingegeben haben. Das Konzept Datenbewusstsein ist recht neu und basiert auf sehr aktuellen Entwicklungen, weshalb es sich sicherlich weiter ausdifferenzieren und in unterschiedlichen (etwa schulischen) Kontexten bewähren muss. Mit dem Konzept zeigen wir jedoch auf, in welche Richtung fachdidaktische Forschung in näherer Zukunft verstärkt Innovationen hervorbringen wird und neuere fachwissenschaftlich und gesellschaftlich relevante Fragen aufnehmen kann.

## Literatur

- [Be15] Berendt, B. et al.: Datenschutz im 21. Jahrhundert: Ist Schutz der Privatsphäre (noch) möglich? In (Gallenbacher, J., Hrsg.): INFOS 2015. Bonn, S. 33–42, 2015.
- [BK15] Bode, M.; Kristensen, D.: The digital doppelgänger within. A study on self-tracking and the quantified self-movement. In (Canniford, R.; Bajde, D., Hrsg.): *Assembling Consumption*. Routledge, 2015.
- [Bu16] Burrell, J.: How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms. *Big Data & Society* 3/1, S. 1–12, 2016.
- [Ga02] Gal, I.: Adults’ Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review* 70/1, S. 1–25, 2002.
- [LM20] Long, D.; Magerko, B.: What is AI Literacy? Competencies and Design Considerations. In: 2020 CHI. CHI ’20, ACM, Honolulu, HI, USA, S. 1–16, 2020.
- [Ma17] Mau, S.: *Das metrische Wir: Über die Quantifizierung des Sozialen*. Suhrkamp Verlag, Berlin, 2017.
- [OE14] OECD: Summary of the OECD Privacy Expert Roundtable „Protecting Privacy in a Data-driven Economy: Taking Stock of Current Thinking“. DSTI/ICCP/REG(2014)3, 2014.
- [PS19] Pangrazio, L.; Selwyn, N.: ‘Personal data literacies’: A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society* 21/2, S. 419–437, 2019.
- [Ra19] Rahwan, I. et al.: Machine behaviour. *Nature* 568/7753, S. 477–486, 2019.
- [Ri15] Ridsdale, C. et al.: *Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report*, 2015.
- [RK20] Register, Y.; Ko, A. J.: Learning Machine Learning with Personal Data Helps Stakeholders Ground Advocacy Arguments in Model Mechanics. In: ICER ’20. ACM, Virtual Event New Zealand, S. 67–78, 10. Aug. 2020.
- [SB18] Schulte, C.; Budde, L.: A Framework for Computing Education: Hybrid Interaction System: The Need for a Bigger Picture in Computing Education. In: *Koli Calling ’18*. Bd. 18, ACM, Koli, Finland, S. 10, 2018.
- [SFN18] Shapiro, R. B.; Fiebrink, R.; Norvig, P.: How machine learning impacts the undergraduate computing curriculum. *CACM* 61/11, S. 27–29, 2018.
- [SS18] Smith, B.; Shum, H.: *The Future Computed: Artificial Intelligence and Its Role in Society*. Microsoft, Redmond, 2018.
- [Te20] Tedre, M. et al.: Machine Learning Introduces New Perspectives to Data Agency in K-12 Computing Education. In: 2020 FIE. IEEE, S. 1–8, 2020.
- [Zu18] Zuboff, S.: *Das Zeitalter des Überwachungskapitalismus*. Campus Verlag, Frankfurt am Main, 2018.