

# Von Datenmanagement zu Data Literacy: Grundlagen und Entwicklung eines Data-Literacy-Kompetenzmodells

Andreas Grillenberger  
Freie Universität Berlin  
Didaktik der Informatik

E-Mail-Adresse: andreas.grillenberger@fu-berlin.de

## Zusammenfassung

„Daten stellen einen wichtigen Rohstoff des 21. Jahrhunderts dar“ – solche und ähnliche Einschätzungen betonen die Relevanz von Daten für unsere Gesellschaft. Obwohl heute nahezu jede Person kontinuierlich umfangreiche Daten produziert, bleiben sowohl die Herausforderungen und Gefahren als auch die Möglichkeiten, die damit einhergehen, oft eher im Verborgenen. Für einen fundierten und souveränen Umgang mit Daten sind Kompetenzen im Bereich der Arbeit mit und Verwaltung von Daten jedoch unverzichtbar. Im Beitrag wird daher ein grundlegender Überblick über diesen Themenbereich gegeben, die Bedeutung von Datenkompetenzen in verschiedenen Bereichen herausgearbeitet und das Thema aus Perspektive der Informatikdidaktik detaillierter betrachtet. Dazu werden zwei Modelle vorgestellt, die das Thema nicht nur aus informatischer und informatikdidaktischer Sicht, sondern auch die interdisziplinäre Diskussion um Datenkompetenzen beflügeln und fundieren können, wie erste Erfahrungen zeigen. Zuletzt wird ein Einblick in die Förderung von Datenkompetenzen im Bereich der Schulinformatik gegeben, die gegebenenfalls auch als prototypische Ideen für die Berücksichtigung von Datenkompetenzen im interdisziplinären Bereich dienen können.

## 1. Einleitung

Mit der zunehmenden Digitalisierung von Geräten, Diensten und Prozessen in unserer heutigen Lebenswelt geht auch eine zunehmende Bedeutung von Daten einher. Diese sind heute allgegenwärtig und werden in allen Lebenssituationen erzeugt. Dadurch entstehen für alle Personen vielfältige Herausforderungen, zu deren Beherrschung gewisse Kompetenzen aus dem Bereich *Daten* nötig sind. Diese *grundlegenden Datenkompetenzen* werden oft auch als *Data Literacy* bezeichnet und häufig als Teil der heute weitestgehend als notwendig erachteten *digitalen Kompetenzen* betrachtet (vgl. z. B. Vuorikari, Punie, Carretero Gomez, & Van Den Brande, 2016).

Die Notwendigkeit von grundlegenden Datenkompetenzen für alle Personen lässt sich insbesondere auf drei zentrale Argumentationsbereiche zurückführen:

- 1) *Gesellschaftlich-kulturell*: Während Daten bis vor wenigen Jahren speziell als Thema der Informatik betrachtet wurden, die sich deren Darstellung und Verarbeitung verschreibt (vgl. z. B. Claus & Schwill, 2006), werden sie heute – beabsichtigt oder unbeabsichtigt – durch alle Personen erzeugt und durch vielfältige Prozesse verarbeitet. Dies umfasst die Erschließung neuen Wissens

genauso wie die Bewertung bzw. Einstufung von Personen anhand von Daten. Für die von einer konkreten Datennutzung betroffenen Personen ist die Datensammlung zwar oft prinzipiell wahrnehmbar, ohne entsprechende Kompetenzen sind aber weder die damit einhergehenden Gefahren noch Möglichkeiten klar erkennbar. Entsprechend wird meist eine eher passive datenproduzierende Rolle eingenommen. Die Ermächtigung von Personen zur Arbeit mit Daten ist daher ein zentraler Aspekt der Data Literacy, ohne dass dabei auf die vertiefte Ausbildung fachlicher Kompetenzen, wie sie beispielsweise für einen Data Scientist<sup>1</sup> notwendig wären, abgezielt wird.

- 2) *Wissenschaftspropädeutisch*: Die Verwaltung von und Arbeit mit Daten stellt ein Themenfeld der Informatik dar, das sich unter anderem in der aktuellen Forschung zu Themen wie *Big Data*, *Data Mining* und *Datenstromsystemen* äußert (GI, 2018), aber auch in tradierten Themen wie *Datenbanken*. Daten stellen jedoch auch einen Gegenstandsbereich dar, der aus Gründen der Wissenschaftspropädeutik mindestens im Informatikunterricht eine wichtige Rolle spielen sollte, um einen umfassenden Eindruck der Informatik zu vermitteln und deren Breite zu verdeutlichen, die weit über Bereiche wie *Algorithmik* und *Programmierung* hinausgeht. Dies ist jedoch bislang nur sehr eingeschränkt der Fall (Grillenberger & Romeike, 2014).
- 3) *Interdisziplinär*: Daten werden heute auch für informatikferne Fächer und Berufsbilder zunehmend relevant und prägen das jeweilige Fachverständnis insbesondere in der Forschung (GI, 2018; Hey, Tansley, & Tolle, 2009), aber auch in eher anwendungsorientierten Bereichen. Beispielsweise sind heute in der Physik und der Astronomie Daten und deren Verwaltung und Analyse zentral, wenn Teilchenbeschleuniger oder Radioteleskope mehrere Terra- oder sogar Petabyte an Daten erzeugen, die ausgewertet werden müssen bzw. aus denen zum Teil nur noch Ausschnitte zur Analyse ausgewählt werden können. Auch in Geisteswissenschaften sind deutliche Veränderungen des Forschungsprozesses durch neue Möglichkeiten zur Erfassung und Nutzung großer und komplexer<sup>2</sup> Datenmengen erkennbar, die sich beispielsweise in den sog. *Digital Humanities* widerspiegeln.

Um einen fachlich adäquaten Blick auf die Arbeit mit Daten zu ermöglichen, sollte die Betrachtung der aus Perspektive dieser Argumentationsbereiche als notwendig erachteten Data-Literacy-Kompetenzen aus verschiedenen Blickwinkeln erfolgen, die sich aus den der Arbeit mit Daten inhärenten fachlichen Bezügen ergeben:

- Die *mathematische und statistische Perspektive* auf Daten erlaubt beispielsweise Aussagen über die Aussagekraft von Daten und auf diesen basierenden Analysen, aber gibt auch einen Einblick in die Fundierung verschiedener Analysemethoden und in deren zugrundeliegenden Prinzipien.

---

<sup>1</sup> Data Science bezeichnet eine interdisziplinäre Fachrichtung, die Aspekte von Informatik, Mathematik und Anwendungswissenschaften einbezieht und darauf abzielt, Personen zur systematischen Arbeit mit Daten, oft im Unternehmenskontext, zu befähigen (GI, 2018).

<sup>2</sup> Im Allgemeinen muss, wenn von Big Data gesprochen wird, nicht nur die Größe des Datensatzes, sondern auch dessen Komplexität berücksichtigt werden, da dieser oft aus verschiedenen Arten zueinander in Beziehung stehender Daten besteht.

- Die *informatische Perspektive* gibt Einblick in die dynamischen Aspekte der automatisierten Verarbeitung von Daten und die zugrundeliegenden Ideen und Prinzipien, betrachtet aber auch die statischen Aspekte, wie Strukturierung, Speicherung und Verwaltung von Daten mit Informatiksystemen.
- Die *anwendungsorientierte Perspektive* bildet den sinnstiftenden Kontext für die Datennutzung, verleiht den angestrebten Datenkompetenzen einen Sinn und stellt einen Bezug zur Realwelt her. Als Anwendungsfach können, entsprechend der Interdisziplinarität der Kompetenzen, Themen aus vielfältigen Fächern herangezogen werden.

In diesem Beitrag wird die Annäherung an das umfangreiche Themenfeld der Datenkompetenzen aus Sicht der Informatik und der Informatikdidaktik beschrieben. Dazu wird zuerst das diesem Beitrag zugrundeliegende Begriffsverständnis der *Data Literacy* charakterisiert und diese von verwandten und überschneidenden Konzepten abgegrenzt. Daraufhin wird erläutert, wie zur Aufarbeitung des Themenkomplexes aus informatikdidaktischer Perspektive verfahren wurde. Dabei wurde zuerst eine fachliche Perspektive eingenommen, um einen Einblick in das Datenkompetenzen aus fachlicher Sicht zugrundeliegende Gebiet *Datenmanagement* zu erlangen. Diese Betrachtung mündete in ein Modell der Schlüsselkonzepte dieses Fachgebiets. Eine eher auf allgemeinbildende Aspekte gerichtete Perspektive wurde daraufhin eingenommen, indem ein theoretisch fundiertes *Data-Literacy-Kompetenzmodell* entwickelt wurde, das, trotz seiner Genese aus Perspektive der Informatikdidaktik, nicht nur informatische Aspekte miteinschließt, sondern durch eine weitergehende Unterfütterung durch mathematische und anwendungsorientierte Aspekte auch eine interdisziplinäre Betrachtungsweise erlaubt. Dazu wird ein Ausblick auf den zukünftigen Einsatz der Modelle als Basis für die Förderung von Data-Literacy-Kompetenzen im interdisziplinären Hochschulkontext gegeben. Zuletzt werden zwei Ansätze zur Förderung von in diesem Modell fundierten Kompetenzen vorgestellt und entsprechende Erfahrungen skizziert.

## 2. Begriffsverständnis *Data Literacy*

Der Begriff *Data Literacy* wird in diesem Beitrag im Sinne der Definition des Begriffs durch Ridsdale et al. (2015) als Fähigkeit verstanden, „*Daten auf kritische Art und Weise zu sammeln, zu managen, zu bewerten und anzuwenden*“<sup>3</sup>. Dieses relativ umfassende Verständnis wird national wie international häufig zugrunde gelegt, bspw. wenn es um Projekte zur Förderung von solchen Kompetenzen geht.

Entsprechend der obigen Definition weist die *Data Literacy* Überschneidungen zu weiteren Konzepten auf, die sich insbesondere hinsichtlich ihrer jeweiligen Zielsetzungen unterscheiden:

*Statistical Literacy* setzt den Schwerpunkt auf statistische Aspekte (Schild, 2018) und überschneidet sich daher relativ deutlich mit *Data Literacy*. Gleichzeitig kann diese jedoch nicht als Teilmenge der *Data Literacy* aufgefasst werden, da die statistische Perspektive weitere Aspekte ergänzt. Zum Teil wird in diesen Zusammenhängen auch von *Statistical Thinking* gesprochen.

---

<sup>3</sup> Übersetzung durch Hochschulforum Digitalisierung, (o. J.).

Das informatische Äquivalent stellt das *Computational Thinking* dar, das z. B. im *Barefoot-Modell* (Computing at School, o. J.) charakterisiert wird. Wie auch die Statistical Literacy weist es Überschneidungen mit Data Literacy auf. Diese geht wegen ihrer Interdisziplinarität aber auch über Computational Thinking hinaus.

Auch die *Information Literacy* weist nach Heidrich, Bauer, & Krupka (2018) starke Überschneidung mit Data Literacy auf, da sie nach Carlson & Johnston (2015) als „Fähigkeit, Informationen aus verschiedenen Formaten zu finden, zu managen und zu verwenden“<sup>4</sup> verstanden wird. Dabei steht jedoch insbesondere die Nutzung und Aufbereitung vorhandener Informationen im Vordergrund (ACRL Board, 2016).

Während die bislang vorgestellten Konzepte Überschneidungen mit der Data Literacy aufweisen, kann die *Digital Literacy* (Vuorikari et al., 2016) als übergeordneter Begriff verstanden werden, der große Teile der Data Literacy berücksichtigt – da diese jedoch nicht auf die digitale Arbeit mit Daten beschränkt ist, jedoch nicht vollständig.

### 3. Data Literacy aus Perspektive der Informatik

Das informatische Fundament für grundlegende Datenkompetenzen ist insbesondere im Fachgebiet *Datenmanagement* zu finden:

„Datenmanagement umfasst die informatischen Grundlagen des Umgangs mit und der Verwaltung und Verarbeitung von Daten. Es bezieht alle Phasen des Datenlebenszyklus [...] mit ein.“ (Grillenberger, 2019)

Das Fachgebiet Datenmanagement ist dabei noch relativ jung, umfasst jedoch auch Themen aus dem Bereich Datenbanken, die bereits eine für die Informatik lange andauernde Tradition aufweisen. Um aus informatischer Perspektive einen Einblick in die Data Literacy zu erlangen und die aus dieser Sicht zentralen Kompetenzen herauszuarbeiten, musste zuerst eine ausreichende Aufarbeitung dieses Fachgebiets stattfinden. Dazu wurden, basierend auf einer explorativen Analyse und in Anlehnung an das Modell der *Great Principles of Computing* (Denning, 2003), die diesem Fachgebiet zugrundeliegenden Praktiken, Kerntechnologien, Mechanismen und Entwurfsprinzipien herausgearbeitet und in einem Modell der *Schlüsselkonzepte des Datenmanagements* zusammengefasst (vgl. Abb. 1; Grillenberger & Romeike, 2017). Dieses Modell vereinfacht den Einblick in das umfangreiche Fachgebiet, verdeutlicht verschiedene Perspektiven auf dieses, indem es nicht nur die Ideen hinter



Abbildung 1 Modell der Schlüsselkonzepte des Datenmanagements



Abbildung 2 Praktiken des Datenmanagements, visualisiert als Datenlebenszyklusmodell

<sup>4</sup> Übersetzung durch Heidrich et al., (2018).

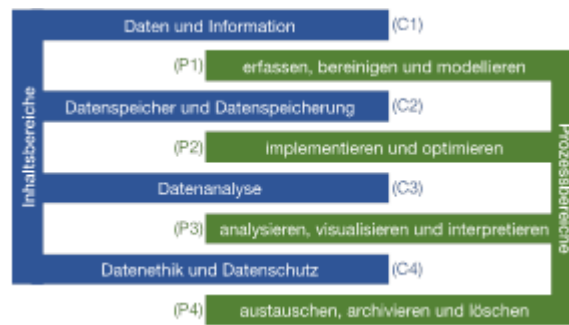
Datenmanagementsystemen beleuchtet, sondern auch eine technologische und eine praktische Perspektive einnimmt. Die dargestellten Praktiken können auch als Datenlebenszyklus betrachtet werden (vgl. Abb. 2) und geben somit eine wichtige Orientierung für die Arbeit mit Daten. Somit erlaubt das Modell aus fachlicher Sicht einen Einblick in Grundlagen der Datenverwaltung und -verarbeitung, die als Basis für grundlegende Datenkompetenzen dienen können. Da diese jedoch nicht rein aus Perspektive der Informatik geprägt sind und da dem Modell noch keine didaktische Reduktion bzw. Auswahl zugrunde liegt, können diese nicht direkt aus dem Modell abgeleitet werden, sodass eine entsprechende Weiterentwicklung nötig wurde.

Diese Weiterentwicklung fand unter Berücksichtigung der durch das Modell der Schlüsselkonzepte explizierten fachlichen Grundlagen und unter Einbeziehung einer allgemeinbildenden Perspektive, die für eine Data Literacy im angestrebten Sinne zentral ist, statt. Um Lehrpersonen, insbesondere für den allgemeinbildenden Unterricht, eine Orientierung und Grundlage zu geben, wurde in diesem Schritt ein *Data-Literacy-Kompetenzmodell* entwickelt. Das Modell orientiert sich hinsichtlich seiner Struktur an dem in der Schulinformatik bewährten Kompetenzmodell, das den *Empfehlungen für Bildungsstandards in Informatik* (GI, 2008) zugrunde liegt. Diese sehen eine Unterteilung der Kompetenzbereiche in zwei Typen vor: *Inhaltsbereiche* berücksichtigen die jeweils zugrundeliegenden zentralen fachlichen Inhalte, Ideen und Konzepte. Sie stellen somit die inhaltliche Dimension der Kompetenzen dar. *Prozessbereiche* befassen sich mit praktischen Tätigkeiten, die beim Umgang und der Arbeit mit dem jeweiligen Thema zentral sind und einen praktischen Zugang eröffnen. Beide Arten von Kompetenzbereichen sind stark miteinander verwoben: *„Inhalte und Prozesse [sind] aufeinander angewiesen. Die Prozesskompetenzen werden an der Arbeit mit den Inhalten erworben und die Inhalte stünden ohne Prozesse in der Gefahr, zu einer Wissenssammlung zu verkommen.“* (GI, 2008). Daher sind alle Kompetenzen jeweils einem Prozess- und Inhaltsbereich zugeordnet.

Das Data-Literacy-Kompetenzmodell wurde aus den fachlichen Grundlagen abgeleitet, die sich im Modell der Schlüsselkonzepte des Datenmanagements sowie den fachlichen Inhalten der verwandten Data Science<sup>5</sup> widerspiegeln. Dazu wurden, in Zusammenarbeit mit Lehrpersonen mit unterschiedlichem Hintergrundwissen, Schwierigkeiten bei der Adaption des Themas für den Unterricht und Anforderungen an das Modell herausgearbeitet. Diese wurden in die Entwicklung einbezogen, z. B. durch Zusammenführung von Prozessbereichen, die im Unterricht kaum trennbar sind, wie „analysieren, visualisieren und interpretieren“. Durch die fachliche Fundierung kann jedoch weiterhin ein tiefergehender Einblick gewonnen werden, sodass Details nicht vernachlässigt, sondern nur im konkreten Modell ausgeblendet wurden. Einen detaillierteren Einblick in die Entwicklung des in Abb. 3 dargestellten Kompetenzmodells geben Grillenberger & Romeike (2018) und Grillenberger (2019).

---

<sup>5</sup> Im Gegensatz zur Data Literacy, die grundlegende und ggf. für jede Person notwendige Aspekte des Umgangs mit Daten berücksichtigt, legt Data Science den Fokus auf die Ausbildung detaillierten Professionswissens.



**Abbildung 3** Data-Literacy-Kompetenzmodell

In der vorliegenden Form dient das Data-Literacy-Kompetenzmodell primär als erste Orientierung für die Förderung von Datenkompetenzen und betont die Relevanz sowohl praktischer Zugänge als auch fachlich fundierter Inhalte. Eine detaillierte Ausgestaltung des Modells steht jedoch noch aus, u. a. durch Identifikation und Validierung konkreter Kompetenzen, durch Einführung von Kompetenzniveaus/-stufen und von Möglichkeiten zur Überprüfung des Kompetenzerwerbs. Auch die Weiterentwicklung des Modells unter Berücksichtigung überfachlicher Aspekte, die sich nach aktuellen Diskussionen vermutlich an vielen Stellen verorten lassen (vgl. unten), aber ggf. auch Ergänzungen und Veränderungen nach sich ziehen, ist ein wichtiger Gegenstand der aktuellen und zukünftigen Arbeit.

#### 4. Interdisziplinäre Förderung von Datenkompetenzen

Wie bereits dargestellt, ist die Förderung von Data Literacy zwar aufgrund fachlicher Bezüge im Interesse der Informatik, aber nicht auf diese beschränkt: Sinnstiftende Arbeit mit Daten findet immer in einem Kontext statt, der nicht nur aus der Informatik stammen kann, sondern oft aus einem anderen Fach. Bestrebungen zur Förderung von Data Literacy zeigen sich daher auch im interdisziplinären Bereich: Dies ist beispielsweise in Projekten wie *Daten Lesen Lernen* (Universität Göttingen, 2018) erkennbar, die eine Förderung von Data-Literacy-Kompetenzen als interdisziplinäre Aufgabe ansehen und verschiedene Fächer einbeziehen. Aus diesem Grund wurden erste Überlegungen und Diskussionen angestellt, die beiden vorgestellten Modelle als Basis für die interdisziplinäre Förderung von Data Literacy zu verwenden. In Diskussionen dazu zeigte sich, dass die Vertreter der beteiligten Fachrichtungen (u. a. Biologie, Geographie, Sprachen) in den Modellen die Herausforderungen und Anforderungen, die sich in den jeweiligen Fächern ergeben, wiedererkennen: Dabei schien besonders zentral, dass die Praktiken des Datenmanagements, die sich in den Prozessbereichen des Kompetenzmodells widerspiegeln, als kompletter Prozess aufgefasst werden (Abb. 2) und damit als Orientierung für die Arbeit mit Daten dienen können. Gleichzeitig war eindeutig feststellbar, dass das Kompetenzmodell als Orientierung nicht ausreicht, da dieses relativ viele Informationen ausblendet, sodass die Datenmanagement-Schlüsselkonzepte als notwendige Basis erachtet und von den Fächern akzeptiert wurden. Weiterhin zeigte sich bereits in dieser frühen Phase der Planung einer interdisziplinären Förderung von Datenkompetenzen eine zentrale Herausforderung, die den gesamten Prozess begleiten wird – das unterschiedliche Begriffsverständnis: So ist beispielsweise der Begriff (Daten-)Modellierung aus informatischer Sicht klar belegt, in anderen Fächern wird aber ein teils deutlich anders ausgeprägter Modellbegriff genutzt, sodass es notwendig sein wird, selbst augenscheinlich klar verständliche Begriffe klar zu charakterisieren, um

Missverständnisse zu vermeiden – ein gemeinsames Glossar der Begriffe ist daher unabdingbar für eine interdisziplinäre Zusammenarbeit in diesem Gebiet.

Basierend auf den hier nur knapp geschilderten ersten Erfahrungen, ist eine interdisziplinäre Förderung von Datenkompetenzen aus Perspektive verschiedener Fächer ein wichtiges Ziel. Die zuvor vorgestellten Modelle sind dabei als Basis für eine gemeinsame Arbeit geeignet, müssen aber mit den Fächern ausgestaltet und ergänzt werden, um bspw. durch weitere Inhaltsbereiche Fachspezifika abzudecken.

## 5. Praktische Beispiele

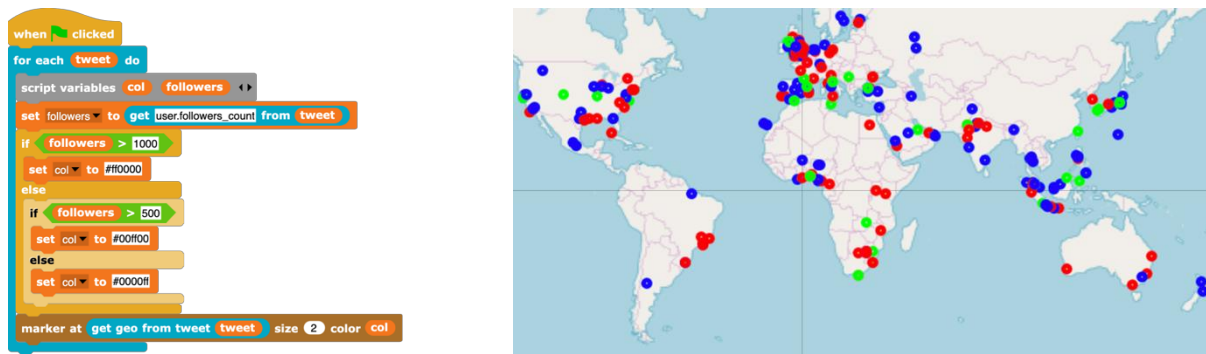
Auch zur praktischen Förderung von Data Literacy konnten Erfahrungen gesammelt werden. Die beiden beschriebenen Beispiele wurden für den Informatikunterricht konzipiert, können aber auch auf andere Bildungsniveaus und -kontexte übertragen werden. Beide Beispiele befassen sich mit der Datenanalyse und verbinden den Prozessbereich P3 (*Analysieren, Visualisieren, Interpretieren*) mit den Inhaltsbereichen C1 (*Information und Daten*), C3 (*Datenanalyse*) und C4 (*Datenethik und Datenschutz*). Aus inhaltlicher Sicht stehen Aspekte wie die Unterscheidung von Information und Daten, die Aussagekraft von Daten und die Bedeutung von Metadaten (C1), ein Verständnis grundlegender Analysemethoden (C3) und ethische Fragestellungen und Herausforderungen (C4) im Vordergrund. Als praktischer Zugang werden einfach nutzbare Werkzeuge, die wenige Vorkenntnisse benötigen, zum Teil aber auch die händische Datenanalyse, gewählt. Diese werden durch Visualisierungen und die Interpretation und kritische Reflektion der Ergebnisse und der zugrundeliegenden Daten (P3) ergänzt. Weitere Details zu beiden Beispielen finden sich in Grillenberger (2019).

### Beispiel 1: Analyse des Twitter-Datenstroms mit Snap!Twitter

Im ersten Beispiel sollte Schülerinnen und Schülern ein einfacher Zugang zum Thema Datenanalyse eröffnet werden, ohne spezifische Vorkenntnisse hinsichtlich Programmierung, Datenanalyse oder -verwaltung vorauszusetzen. Dabei sollten die Lernenden erste eigene praktische Erfahrungen in diesem Bereich sammeln und grundlegende Kenntnisse über die Analyseideen *Klassifikation* und *Clusterbildung* erwerben. Als Werkzeug wurde die blockbasierte Programmierumgebung *Snap!* gewählt und so erweitert, dass ein Zugriff auf den Twitter-Datenstrom ermöglicht wird, da dieser eine umfangreiche Datenquelle darstellt, die sich durch ihre Alltagsnähe, die umfangreichen (Meta-)Daten und eine gute Zugänglichkeit auszeichnet. Aufgrund von Einschränkungen der Twitter-API kann nur auf ca. 40-60 Tweets pro Sekunde zugegriffen werden. Mit über 150 verfügbaren Metaattributen sind trotzdem spannende Datenanalysen möglich, sodass die Datenmenge für das angestrebte Ziel ausreicht. Da viele Tweets auch den Nutzerstandort beinhalten, wurde *Snap!* weiterhin so erweitert, dass Kartendarstellungen und Diagramme erstellt werden können, sodass ein visueller Zugang zur Datenanalyse ermöglicht wird.

In Abb. 4 ist ein Analysebeispiel dargestellt, das in einer Schülergruppe entstand, die unterschiedliche Beteiligung im sozialen Netzwerk je nach Region visualisierte. Dazu





Die Farben visualisieren die Followerzahl des Autors (rot  $\hat{=}$  >1000; blau  $\hat{=}$   $\leq$  500; grün dazwischen). Kartendaten: © OpenStreetMap

**Abbildung 4** Code und Kartendarstellung einer Analyse mit Snap!Twitter

wurden Tweets auf einer Karte dargestellt, je nach Followeranzahl des Autors eingefärbt und durch manuelles, visuelles Clustering diese Daten interpretiert: Im Beispiel ist ein beträchtlicher Teil der roten und grünen Tweets (über 1000 bzw. 500 Follower) in Europa angesiedelt. Dies kann z. B. die Interpretation nahelegen, dass soziale Netzwerke in Europa eine besonders starke Bedeutung haben. Gleichzeitig ist aber auch erkennbar, dass sie z. B. in Russland unbedeutend scheinen – was eine Diskussion über die Aussagekraft der Daten und der Analyse nahelegt. Erste Erfahrungen zeigen, dass durch das Werkzeug ein interessanter, motivierender Zugang eröffnet wurde, der zu ersten Diskussionen zur Aussagekraft von Daten (z. B. Bias durch den Analysezeitpunkt) oder über Privatsphäre hinleitet und somit als Sensibilisierung aber auch als Basis für eine vertiefte Thematisierung dient.

## Beispiel 2: Vorhersage von Schulnoten anhand von Schülerdaten

In einem zweiten Beispiel wurden nicht nur Daten analysiert, sondern diese auch als Basis für Vorhersagen verwendet. Dazu wurden Grundlagen wie Klassifikation und Klassifikationsbäume anhand händischer Analysebeispiele und kleiner fiktiver Datensätze eingeführt, sodass die Schülerinnen und Schüler einen Einblick in den Analyseprozess und die Entstehung von Vorhersagen erhielten. Da händisch durchgeführte Datenanalysen jedoch kaum die Mächtigkeit von Analysen verdeutlichen können, wurde auf Basis dieser Grundlage ein echter und größerer Datensatz analysiert, damit die Schülerinnen und Schüler auch erkennen, dass sie selbst echte Daten analysieren können. Dazu wurde mit dem professionellen aber einfach nutzbaren Datenanalysewerkzeug *Orange* (vgl. Abb. 5) eine Datenanalyse durchgeführt. Durch eine am Datenfluss orientierte Darstellung eröffnet das Werkzeug einen relativ einfachen Zugang, der für die Lernenden (9. Klasse einer Realschule) keine Herausforderung darstellte. Bei der Analyse wurde aus Gründen der Komplexität nicht im Detail untersucht, wie die beispielsweise zur Erstellung des Klassifikationsbaums genutzten Algorithmen funktionieren, sondern auf das vorher händisch erworbene prinzipielle Verständnis gesetzt, sodass diese trotzdem nicht als komplette Blackbox erschienen.

Um auch hier einen motivierenden Zugang zu nutzen, der die Schülerinnen und Schüler direkt betrifft, wurde ein Datensatz mit Daten von über 600 portugiesischen Schülerinnen und Schülern gewählt (Cortez & Silva, 2008), der nicht nur drei Noten von diesen, sondern auch Informationen über deren Lebensverhältnisse beinhaltet. Anhand der Daten wurde ein Modell zur Vorhersage der dritten Schulnote anhand



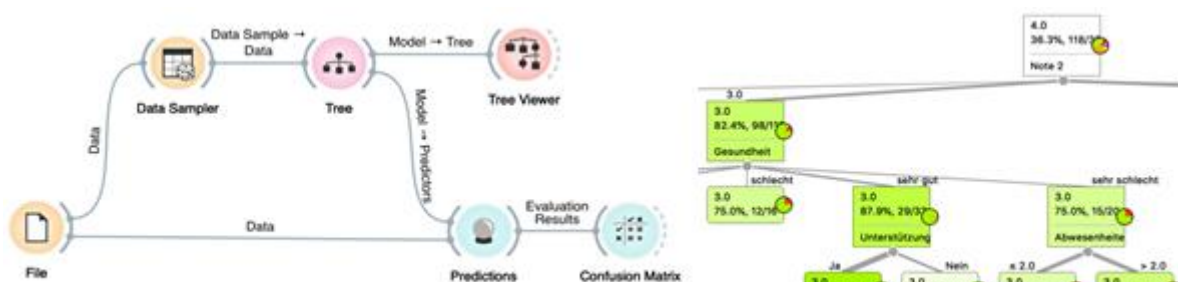
aller anderen Daten erstellt. Dieses Modell lieferte erstaunlich genaue Ergebnisse, was eine entsprechende Diskussion anregte: Auf dieser Basis konnten somit nicht nur Gefahren einer automatisierten Einstufung bzw. Bewertung von Personen erkannt und abgewägt werden, sondern auch ethische Aspekte (wie Stigmatisierung) diskutiert und ein grundsätzliches Bewusstsein dafür geschaffen werden.

## 6. Zusammenfassung und Fazit

In diesem Beitrag wurde, basierend auf einer Argumentation der Notwendigkeit von grundlegenden Datenkompetenzen im Sinne einer Data Literacy, ein umfassender Überblick über diese und das aus informatischer Perspektive zugrundeliegende Datenmanagement gegeben. Bei der Entwicklung der beiden vorgestellten Modelle wurde primär die Perspektive der Schulinformatik eingenommen, aus der auch die vorgestellten Beispiele zur praktischen Umsetzung im Rahmen der informatischen Bildung stammen. Erste Eindrücke eines interdisziplinären Diskurses zeigen jedoch, dass das fachlich orientierte Modell der Schlüsselkonzepte des Datenmanagements sowie das darauf basierende Kompetenzmodell der Data Literacy, als Basis für die anlaufende interdisziplinäre Förderung von Datenkompetenzen dienen und somit auch im Hochschulkontext genutzt werden können: Basierend auf ersten Eindrücken scheinen die aus informatischer Sicht entstandenen Modelle auch in diesem Kontext hilfreich zu sein, sie können mit voraussichtlich nur geringen Anpassungen auch auf die Förderung von Datenkompetenzen in weiteren Fächern übertragen werden. Das in diesem Beitrag aufgeworfene Themenfeld scheint somit nicht mehr nur aus informatischer Sicht und für die Schulbildung wichtig, sondern für alle Bildungsniveaus und auch aus interdisziplinärer Perspektive, sodass die hier vorgestellten Modelle zukünftig noch weiter ausgestaltet und vertieft werden müssen.

## 7. Referenzen

- ACRL Board. (2016). Framework for Information Literacy for Higher Education. Retrieved from <http://www.ala.org/acrl/standards/ilframework>
- Carlson, J., & Johnston, L. R. (Eds.). (2015). *Data-Information Literacy*. Purdue University Press.
- Claus, V., & Schwill, A. (2006). *Duden Informatik A-Z: ein Fachlexikon für Studium, Ausbildung und Beruf*. Mannheim: Dudenverlag.
- Computing at School. (o. J.). Computational Thinking, <https://community.computingschool.org.uk/files/8221/original.pdf> (Abgerufen am 31.05.2020)



**Abbildung 5** Darstellung einer Datenanalyse in Orange und Ausschnitt aus einem Klassifikationsbaum

- Cortez, P., & Silva, A. (2008). Using Data Mining to Predict Secondary School Student Performance. In *Proceedings of 5th Annual Future Business Technology Conference, Porto, 2008*. EUROSIS-ETI.
- Denning, P. J. (2003). Great principles of computing. *Communications of the ACM*, 46(11), 15.
- GI. (2008). Puhlmann, Hermann & Brinda, Torsten & Fothe, Michael & Friedrich, Steffen & Koerber, Bernhard & Röhner, Gerhard & Schulte, Carsten. (2008). Grundsätze und Standards für die Informatik in der Schule. Bildungsstandards Informatik für die Sekundarstufe I. Empfehlungen der Gesellschaft für Informatik e. V. erarbeitet vom Arbeitskreis "Bildungsstandards". LOG IN. 28. -. (150/151).
- GI. (2018). *Data Literacy und Data Science Education: Digitale Kompetenzen in der Hochschulausbildung*. Verfügbar unter: [https://gi.de/fileadmin/GI/Hauptseite/Aktuelles/Aktionen/Data\\_Literacy/GI\\_DataScience\\_2018-04-20\\_FINAL.pdf](https://gi.de/fileadmin/GI/Hauptseite/Aktuelles/Aktionen/Data_Literacy/GI_DataScience_2018-04-20_FINAL.pdf) (Abgerufen am 17.02.2020)
- Grillenberger, A. (2019). *Von Datenmanagement zu Data Literacy: Informatikdidaktische Aufarbeitung des Gegenstandsbereichs Daten für den allgemeinbildenden Schulunterricht*. Freie Universität Berlin.
- Grillenberger, A., & Romeike, R. (2014). A comparison of the field data management and its representation in secondary CS curricula. In *Proceedings of the 9th Workshop in Primary and Secondary Computing Education (WiPSCE '14)*. Association for Computing Machinery, New York, NY, USA, 29-36. doi: 10.1145/2670757.2670779
- Grillenberger, A., & Romeike, R. (2017). Key concepts of data management: an empirical approach. In Suero Montero, Calkin; Joy, Mike (Eds.), *Proceedings of the 17th Koli Calling Conference on Computing Education Research*. Koli, FI: New York: ACM. doi: [10.1145/3141880.3141886](https://doi.org/10.1145/3141880.3141886)
- Grillenberger, A., & Romeike, R. (2018). Developing a theoretically founded data literacy competency model. In *Proceedings of the 13th Workshop in Primary and Secondary Computing Education (WiPSCE '18)*, October 4-6, 2018, Potsdam, Germany. ACM, New York, NY, USA. doi: 10.1145/3265757.3265766
- Heidrich, J., Bauer, P., & Krupka, D. (2018). *Studie zu übergreifenden Kompetenzen und Studieninhalten in der digitalen Welt am Beispiel von Data Literacy*
- Hey, T., Tansley, S., & Tolle, K. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research. Microsoft Research
- Hochschulforum Digitalisierung. (o. J.). Ausschreibung: "Übergreifende Kompetenzen und Studieninhalte in der digitalen Welt am Beispiel von Data Literacy." Verfügbar unter <https://hochschulforumdigitalisierung.de/de/news/ausschreibung-data-literacy> (Abgerufen am 31.05.2019)
- 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*. Dalhousie University. doi: 10.13140/RG.2.1.1922.5044
- Schild, M. (2018). Information Literacy, Statistical Literacy and Data Literacy. *IASSIST QUARTERLY (IQ)*.
- Universität Göttingen. (2018). Daten Lesen Lernen. Verfügbar unter <http://www.uni-goettingen.de/de/daten+lesen+lernen/592287.html> (Abgerufen am 31.05.2019)
- Vuorikari, R., Punie, Y., Carretero Gomez, S., & Van Den Brande, G. (2016). *DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: the Conceptual Reference Model*. Luxembourg Publication Office of the European Union, doi: 10.2791/11517

## 8. Vita

**Andreas Grillenberger** ist wissenschaftlicher Mitarbeiter der Arbeitsgruppe Didaktik der Informatik der Freien Universität Berlin, an der er 2019 seine Promotion abgeschlossen hat. Im Rahmen seines Dissertationsprojekts erforscht er seit 2013 den Themenkomplex Daten, Datenmanagement und Data Literacy aus Perspektive der Informatikdidaktik und insbesondere der Schulinformatik. Bevor er im Oktober 2018 an die Freie Universität wechselte, war er seit 2013 wissenschaftlicher Mitarbeiter an der Technischen Fakultät und der Professur für Didaktik der Informatik der Friedrich-Alexander-Universität Erlangen-Nürnberg, an der er auch sein Studium für das Lehramt an Gymnasien in den Fächern Informatik und Physik abschloss.