

# Big Data aus Perspektive der Informatikdidaktik

von Andreas Grillenberger

Wie bereits die fachliche Betrachtung von Big Data gezeigt hat, ist die Bedeutung von Big Data für die Informatik heute höher denn je. Die Einflüsse der neuen Möglichkeiten, die sich im Umfeld von Big Data ergeben, sind jedoch nicht auf die Informatik beschränkt: Einerseits entstehen durch Anwendung von Big Data in den völlig unterschiedlichen Fächern auch neue Herausforderungen und Möglichkeiten für diese, beispielsweise indem in der Geografie Geoinformationssysteme (GIS) eine immer höhere Bedeutung erlangen oder die Quellenarbeit im Fach Geschichte durch neue Möglichkeiten ergänzt wird. Auch aus wissenschaftlicher Perspektive eröffnen Big Data und Data Science umfassende neue Möglichkeiten, unter anderem indem mit *Computational Social Science* ein neuer Zweig der Sozialwissenschaften entsteht, der auf diesen Methoden basiert und damit neue Erkenntnisprozesse nutzt (vgl. Scholtes, 2018). Andererseits ist jedoch ebenfalls die gesellschaftliche Bedeutung von Big Data klar hervorzuheben, insbesondere auch im Kontext weiterer Entwicklungen wie der künstlichen Intelligenz und der weiter zunehmenden Automatisierung, die oft starke Zusammenhänge mit der durch Big-Data-Techniken ermöglichten Nutzung immer größerer Daten- und somit Informationsquellen aufweisen und diese umfassend verwenden: Entsprechend existieren Vermutungen, dass Big Data unser Leben mindestens verändert, wenn nicht sogar revolutioniert (vgl. z.B. Mayer-Schönberger/Cukier, 2013).

Durch die vielfältige Bedeutung stellt Big Data somit nicht nur ein Thema für die Fachwissenschaft Informatik dar, sondern auch für andere Fachwissenschaften (meist eher aus Anwenderperspektive), für die Sozialwissenschaften (die sich beispielsweise mit den Folgen der entstehenden neuen Möglichkeiten befassen) und für die Fachdidaktiken – denn Schulunterricht soll auf gesellschaftlich relevante Herausforderungen vorbereiten und Kompetenzen für ein selbstbestimmtes Leben fördern.

In diesem Beitrag wird eine dieser Perspektiven auf das Thema herausgegriffen: die der Informatikdidaktik und des Informatikunterrichts. Auf Basis der Arbeiten des Autors wird der Weg vom fachlichen Themenbereich hin zu einer informatikdidaktischen Betrachtung des Themenfelds skizziert, mit dem Ziel grundlegende Datenkompetenzen zu explizieren, wie sie im Englischen oft unter dem Begriff *Data Literacy* zusammen-

gefasst werden, wobei *Data Literacy* verstanden werden kann als Fähigkeit, „Daten auf kritische Art und Weise zu sammeln, zu managen, zu bewerten und anzuwenden“ (nach Ridsdale/Rothwell u. a., 2015). Der Aufbau dieses Artikels orientiert sich daher am Prozess der Aufbereitung des Themas: Zuerst wird ein Überblick über die dem Themenbereich zugehörigen Aspekte aus fachlicher Perspektive geschaffen, diese dann unter fachdidaktischer Perspektive auf dessen zentrale Aspekte reduziert und zuletzt in ein Kompetenzmodell gegossen, das eine direkte Anwendung im Informatikunterricht erleichtern soll. In der Rubrik *Praxis & Methodik* dieses Hefts werden in weiteren Beiträgen Ideen zur Aufbereitung eines solchen Themenbereichs für den Informatikunterricht vorgestellt.

---

## Zentrale Aspekte

Um das Feld *Big Data* aus informatikdidaktischer Perspektive zu erschließen, muss auch dessen fachlicher Kontext miteinbezogen werden: Big Data bezeichnet oft einen eher modernen, von traditionellen Techniken abweichenden Teilbereich des umfassenderen Fachgebiets *Datenmanagement*. Die Bezeichnung *Datenmanagement* schließt das ursprüngliche Fachgebiet *Datenbanken* der Informatik mit ein, betont jedoch dessen Weiterentwicklung, weg von der Betrachtung nur eines Prinzips der Datenverwaltung hin zu einer Vielfalt von Möglichkeiten, die durch dieses Fachgebiet – unabhängig von der Bezeichnung der Lehrstühle an den Hochschulen oder der Fachbücher – schon seit Langem miteinbezogen werden.

Im Folgenden wird daher nicht nur der Ausschnitt *Big Data* betrachtet, sondern das gesamte Fachgebiet, um eine umfassendere Perspektive zu eröffnen. Wie bereits früher in LOG IN beschrieben (vgl. Grillenberger/Romeike, 2017), bietet sich an dieser Stelle an, das Fachgebiet auf dessen zentralste Konzepte hin zu untersuchen. Auf Basis einer qualitativen Untersuchung des Inhalts diverser Fachbücher und darüber hinausgehender Quellen konnte ein Modell der Schlüsselkonzepte des Datenmanagements entwickelt werden, das einen Überblick über die zentralen Aspekte des Fach-

**Bild 1:**  
**Modell der Schlüsselkonzepte des Datenmanagements.**

gebiets gibt. Bei der Entwicklung des in Bild 1 dargestellten Modells wurde insbesondere berücksichtigt, dass ein solches Thema aus verschiedenen Perspektiven betrachtet werden kann, die allesamt eine Relevanz für den Informatikunterricht haben:



## Kerntechnologien

Die *Kerntechnologien* beschreiben eine technologisch orientierte Perspektive auf das Datenmanagement und umfassen somit verschiedene im Datenmanagement zentrale Systeme. Dabei spiegeln sich auch die technologischen Fortschritte gegenüber dem ursprünglichen Fachgebiet *Datenbanken* wider. Im Gegensatz zu allen anderen Bereichen des Modells strebt der Bereich der Kerntechnologien jedoch nicht zwingend nach einer langfristigen, auf die Zukunft gerichteten Gültigkeit, da diese bei aktuellen technologischen Entwicklungen kaum vorhersagbar ist. Somit umfassen sie neben den klassischen Datenbanken auch modernere Systeme wie dokumentenorientierte Datenspeicher oder Datenstromsysteme (vgl. Grillenberger, 2019b), deren zukünftige Bedeutung derzeit noch nicht sicher ist, aber die die derzeitige Entwicklung im Fachgebiet abbilden.

## Praktiken

Die *Praktiken* des Datenmanagements berücksichtigen die Methoden und Aktivitäten beim Umgang mit Daten und geben somit einen Eindruck davon, welche Möglichkeiten bestehen und welche Tätigkeiten bei einer adäquaten Arbeit mit Daten durchgeführt oder zumindest bedacht werden müssen. Da die Praktiken alle Bereiche des Umgangs mit Daten abdecken, können diese auch als Datenlebenszyklus betrachtet werden (siehe Bild 2) und somit als Leitlinie für einen datenorientierten Informatikunterricht dienen unterstützen, beispielsweise indem sie daran erinnern, oft eher vernachlässigte Aspekte wie die Bereinigung von Daten (z.B. hinsichtlich offensichtlicher Messfehler) oder die Evaluation der Gültigkeit anhand der Daten getroffener Aussagen miteinzubeziehen.

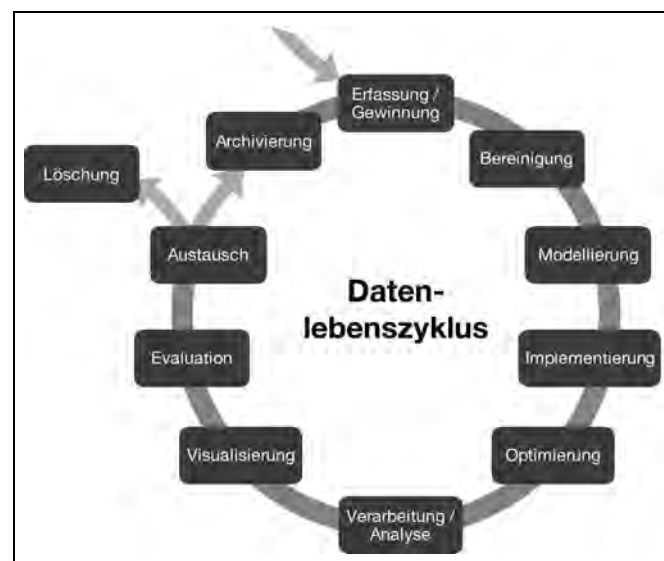
## Entwurfsprinzipien

Die *Entwurfsprinzipien* konzentrieren sich auf zentrale Eigenschaften von Datenmanagementsystemen, die bei deren Entwicklung und Nutzung eine zentrale

Rolle spielen und auch zu deren Charakterisierung herangezogen werden können. Sie umfassen damit beispielsweise die *Konsistenz*, d.h. insbesondere den Grad an Konsistenz, den ein System sicherstellt – denn im Gegensatz zur traditionell meist als gültig erachteten Sichtweise, dass Konsistenz nahezu unverzichtbar ist, verzichten heute immer mehr Systeme auf eine dauerhafte Konsistenz zugunsten einer höheren Geschwindigkeit des Systems. Weiterhin gehören dazu auch Aspekte wie die *Dauerhaftigkeit* des Datenbestands oder der Grad an *Nebenläufigkeit*, den ein System zulässt und unterstützt.

## Mechanismen

Der letzte Bereich des Modells geht am tiefsten in die Funktionsweise von Datenmanagementsystemen hinein und beschreibt die zentralen *Mechanismen*, d.h. jene Konzepte, die die Funktion der Systeme erst ermöglichen. Hierzu gehören also beispielsweise die



**Bild 2: Praktiken des Datenmanagements, aufgefasst als Datenlebenszyklusmodell.**

Strukturierung der Daten, denn ohne gewisse Strukturen kann keine geordnete Datenverwaltung, -nutzung und nicht einmal der Zugriff auf diese stattfinden. Entsprechend werden z.B. Metadaten (wozu natürlich auch einfachste Attribute wie eine künstliche ID eines Datensatzes zählen) eingeführt. Aber auch weitere Aspekte, wie die *Replikation* der Daten zwischen verschiedenen Teilen eines Systems zählen dazu.

### Traditionelle Sicht versus Big-Data-Perspektive

Betrachtet man das Modell aus eher traditioneller Sicht und aus derjenigen von Big Data, so zeigen sich natürlich in den Kerntechnologien deutliche Unterschiede: Beispielsweise kamen durch die neuen Anforderungen von Big Data neue Systeme wie Datenstromsysteme hinzu, andererseits wandelten sich auch traditionelle Systeme immer stärker, sodass nicht mehr nur die seit Langem im Vordergrund stehenden relationalen Datenbanken eine zentrale Rolle spielen, sondern je nach Anwendungszweck auch andere nicht-relationale Datenbankmodelle, die oft durch den Begriff *NoSQL* von den relationalen Datenbanken abgegrenzt werden.

Doch nicht nur in den Kerntechnologien – die sowie einem starken technologischen Wandel unterliegen – können Veränderungen durch Big Data festgestellt werden, sondern insbesondere auch in den Entwurfsprinzipien, die die konzeptionellen Aspekte des Datenmanagements repräsentieren: Während traditionell insbesondere Aspekte wie Datenunabhängigkeit, Integrität, Konsistenz, Isolierung und Dauerhaftigkeit besonders zentral waren, geraten heute vor allem die durch eine verteilte und immer umfassendere Datenverarbeitung relevanten Prinzipien *Verfügbarkeit*, *Partitionstoleranz* und *Nebenläufigkeit* in den Vordergrund. Somit zeigt sich, dass Prinzipien wie die Konsistenz, die jahrzehntlang unverzichtbar schien, heute eher eine Designentscheidung repräsentiert. Man muss sich die Frage stellen, wie wichtig Konsistenz im konkreten Anwendungsfall ist, da diese anderen Prinzipien wie der Verfügbarkeit und der Partitionstoleranz entgegen steht: Soll beispielsweise eine hohe Konsistenz der gespeicherten Daten erreicht werden, sind dafür gewisse Mechanismen nötig, die diese sicherstellen, aber insbesondere in verteilten Datenspeichern unter anderem durch Synchronisation auf verschiedene Server die Geschwindigkeit des Datenspeichers im Gesamten deutlich herabsetzen – und somit der (schnellen) Verfügbarkeit entgegen wirken. Diese Problematik wird übrigens im CAP-Theorem (vgl. Brewer, 2012) verdeutlicht, mit dem aufgezeigt wird, dass die drei Eigenschaften *Konsistenz* (Consistency), *Verfügbarkeit* (Availability) und *Partitions- bzw. Ausfalltoleranz* (Partition Tolerance) nicht gleichzeitig (voll) erreicht werden können.

Was sich jedoch auch durch Big Data nicht verändert hat, sind die Mechanismen, die eine zielgerichtete Verwaltung von Daten ermöglichen. Gerade für den Schulunterricht können dabei diese beiden Bereiche eine wichtige Grundlage darstellen: So sind diverse Mechanismen des Datenmanagements sicherlich in den meisten

Lehrplänen und Bildungsstandards zur Informatik auf die eine oder andere Weise enthalten, wobei die Komplexität der Betrachtung natürlich vom jeweiligen Bildungsniveau abhängt: Beispielsweise wird der grundlegenden Strukturierung von Daten sicherlich mehr Bedeutung beigemessen als der Replikation oder Synchronisation von Daten, wobei gerade bei letzterem die potenziell für den Nutzer sichtbaren und von diesem zu lösenden Synchronisationskonflikte einen interessanten Ansatzpunkt für den Unterricht darstellen können.

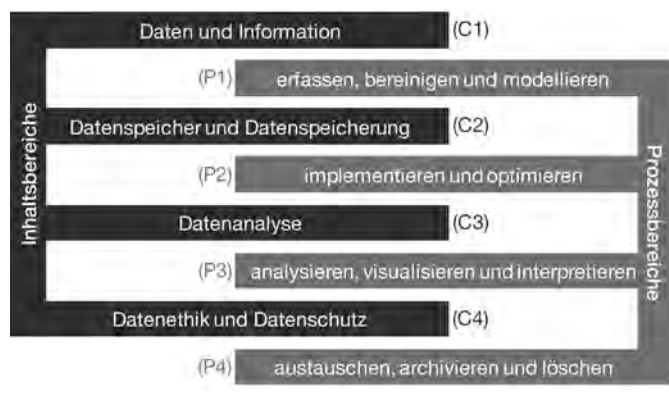
Einen weiteren, eher zeitbeständigen Teil des Modells stellen die Praktiken dar, die – wie zuvor bereits kurz erwähnt – eine besondere Rolle im Informatikunterricht einnehmen können: Durch die Betrachtung als Datenlebenszyklus leisten sie eine wichtige Hilfeleistung sowohl für Lehrende als auch Lernende, indem sie den gesamten Prozess strukturieren und eine Orientierung geben. Dabei muss jedoch beachtet werden, dass verschiedene Datenlebenszyklusmodelle – je nach Herkunft und geplanter Anwendung – unterschiedliche Aspekte dieses Prozesses besonders betonen, im Allgemeinen jedoch sehr große Überschneidungen aufweisen, sodass ihnen eine große Universalität innewohnt. Im Rahmen von Diskussionen mit Lehrkräften konnte jedoch festgestellt werden, dass das zuvor dargestellte Modell – insbesondere aufgrund seiner Detailliertheit – zum Teil Fragen aufwirft, wie die Trennbarkeit der drei Schritte Erfassung/Gewinnung, Bereinigung und Modellierung, die häufig gemeinsam stattfinden dürften. Das ist natürlich eine zulässige Anpassung des Modells, da die Schritte gewisse Überschneidungen aufweisen. Entsprechend wird auch im Folgenden noch mit dem vorgestellten Datenkompetenzmodell eine etwas vereinfachte Betrachtung eingenommen.

Bei der Nutzung des hier vorgestellten Modells der Schlüsselkonzepte des Datenmanagements für den Informatikunterricht stellt sich jedoch weiterhin die Frage, welche Aspekte nun wie und auf welchem Niveau sinnvoll angegangen werden können. Während das Modell zwar einen fachlich orientierten und aus fachdidaktischer Sicht reduzierten Überblick über das Fachgebiet bietet, kann es an dieser Stelle noch nicht deutlich weiterhelfen. Aus diesem Grund wurde nun auf dieser Basis ein weiteres Modell entwickelt, das eine eher für den Unterricht geeignete kompetenzorientierte Betrachtung zugrunde legt.

---

## Von Datenmanagement zu einem Modell grundlegender Datenkompetenzen

Um einen Fokus auf grundlegende Datenkompetenzen einzunehmen, wie sie auch oft als *Data Literacy* bezeichnet werden, kann auf die vorherige fachliche Betrachtung aufgebaut werden. Gleichzeitig bietet sich jedoch auch an, die Struktur des im Informatikunterricht meist genutzten Kompetenzmodells als Basis heranzu-



**Bild 3: Datenkompetenzmodell.**

ziehen: das den Empfehlungen für Informatikstandards der Gesellschaft für Informatik zugrundeliegende. Dieses Modell sieht eine Betrachtung sowohl von Inhaltsbereichen als auch von Prozessbereichen der Informatik vor. Während erstere den Fokus auf die zentralen fachlichen Inhalte, Konzepte und Prinzipien legen, stehen bei Letzteren die praktischen Tätigkeiten im Fokus, die beim Umgang und der Arbeit mit den jeweiligen Themen zentral sind. Inhalts- und Prozessbereiche können jedoch nur im Zusammenspiel miteinander betrachtet werden: „[...] Inhalte und Prozesse [sind] aufeinander angewiesen. Die Prozesskompetenzen werden an der Arbeit mit den Inhalten erworben, ohne die Inhalte wären viele von ihnen nicht spezifisch für die Informatik. Umgekehrt stünden die Inhalte ohne Prozesse in der Gefahr, zu einer Wissenssammlung für Quizshows zu verkommen“ (GI, 2008, S.45). Durch Aufnahme zentraler Prinzipien des Modells der GI-Bildungsstandards soll das entwickelte Datenkompetenzmodell jedoch explizit nicht in Konkurrenz treten, sondern einen fokussierten Blick auf diesen Themenbereich ermöglichen, der dessen Spezifika hervorhebt, jedoch die übergeordneten Aspekte des GI-Modells nicht ausschließen soll.

Um das im Folgenden vorgestellte Modell zu entwickeln, wurde auf die vorher ermittelten und im Modell der Schlüsselkonzepte des Datenmanagements zusammengefassten fachlichen Grundlagen aufgebaut, wobei für die Inhaltsbereiche insbesondere die Entwurfsprinzipien und Mechanismen aus dem Schlüsselkonzeptmodell eine zentrale Rolle spielten, während für die Prozessbereiche die Praktiken bzw. das Datenlebenszyklusmodell grundlegend waren. Zusätzlich wurde durch Befragung von Lehrkräften mit unterschiedlichem Grundwissen hinsichtlich der Probleme beim Verständnis und der Nutzung des Schlüsselkonzeptmodells solche Bereiche identifiziert, die hinsichtlich eines für den Schulunterricht ausgerichteten Kompetenzmodells Verbesserung erfahren müssen. Als Ergebnis konnte ein aus je vier Inhalts- und Prozessbereichen bestehendes Modell entwickelt werden (siehe Bild 3).

Die Inhaltsbereiche des Kompetenzmodells umfassen die fachlichen Grundlagen bzw. inhaltlichen Aspekte. Sie entstanden durch eine Konsolidierung des zuvor

charakterisierten Modells der Schlüsselkonzepte des Datenmanagements hinsichtlich einer Verwendung für den Schulunterricht. Dabei entstanden die folgenden vier Bereiche:

- (C1) *Daten und Information* vereinigt mehrere Aspekte: Er berücksichtigt beispielsweise Unterscheidung von Information und Daten, die unterschiedliche Aussagekraft verschiedener Daten (im Sinne des Informationsgehalts), Methoden zur Erfassung von Information in Form von Daten, aber auch zur Strukturierung von Daten durch Anreicherung mit zusätzlicher Information beispielsweise in Form von Metadaten sowie die Unterscheidung großer und kleiner Datenmengen bezüglich deren Aussagekraft/Informationsgehalt. Entsprechend enthält dieser Inhaltsbereich Aspekte, die umfassende Bezüge zu völlig unterschiedlichen weiteren Bereichen aufweisen, gibt aber insbesondere auch einen Überblick über Daten im Allgemeinen.
- (C2) *Datenspeicher und Datenspeicherung* berücksichtigt grundlegende Mechanismen von Datenspeichern, wie die Repräsentation von Daten bis hin zur physikalischen Ebene, deren Replikation, Synchronisation oder Transport. Gleichzeitig werden aber auch grundlegende Konzepte von Datenspeichern diesem Bereich zugeordnet, wie die Entscheidung, ob Daten konsistent, dauerhaft bzw. hoch verfügbar gespeichert werden sollen. Somit beinhaltet dieser Inhaltsbereich insbesondere die eher statischen Aspekte des Umgangs mit Daten sowie deren Speicherung und die dabei relevanten Konzepte.
- (C3) *Datenanalyse* beschäftigt sich mit allen Aspekten hinsichtlich der Gewinnung von (neuen) Informationen aus Daten – dem, was oft primär unter *Big Data* und *Data Analytics* verstanden wird. Dabei sind insbesondere Methoden, Prinzipien und Algorithmen der Datenanalyse relevant. Dies beinhaltet nicht nur traditionelle Methoden wie die systematische Aggregation bzw. deskriptive Methoden im Allgemeinen, sondern auch explorative, korrelationsbasierte und auf Vorhersagen ausgerichtete Datenanalysen, die oft in starkem Zusammenhang mit maschinellem Lernen stattfinden und Methoden aus diesem Bereich miteinbeziehen. Gerade an dieser Stelle kann auch ein Bezug zu einem weiteren derzeit spannenden Gebiet der Informatik – dem Maschinenlernen – hergestellt werden, da dieses durch entsprechende Methoden die Datenanalyse heute umfassend beeinflusst.
- (C4) *Datenethik und Datenschutz* berücksichtigt alle ethischen, aber auch gesellschaftliche und teils dem Persönlichkeitsschutz entstammende Fragestellungen, die im Rahmen der Arbeit bzw. des Kontakts mit Daten auftreten. Dabei ist hervorzuheben, dass sich Datenschutz hier nicht nur auf den Schutz der Privatsphäre bezieht, sondern auch den Schutz persönlicher Daten beispielsweise durch Maßnahmen wie Verschlüsselung, explizit miteinbezieht.

Im Gegensatz zu den Inhaltsbereichen, denen die Vielfalt von Konzepten, die dem Umgang mit Daten heute innewohnt, legen die Prozessbereiche den Fokus auf den praktischen Umgang mit Daten. Entsprechend weisen sie eine hohe Überschneidung zum vorher vorgestellten Datenlebenszyklus auf, wurden jedoch hinsichtlich der in diesen existierenden Überschneidungen stark konsolidiert. Auch hier entstanden dadurch vier Bereiche:

(P1) *erfassen, bereinigen und modellieren* beschäftigt sich mit den frühen Phasen der Arbeit mit Daten. Mit dem Erfassen, Bereinigen und Modellieren von Daten werden drei Bereiche betrachtet, die nicht voneinander getrennt werden können: Bereits bei der Erfassung ist es notwendig, Daten in einer geeigneten Form zu strukturieren. Damit sind bereits hier zwei Modellierungsentscheidungen nötig: Welcher Ausschnitt der Welt soll in Form von Daten erfasst werden und wie werden diese strukturiert? Gleichzeitig muss bereits zu diesem Zeitpunkt eine Überprüfung der Datenqualität in der Hinsicht stattfinden, dass diese von Fehlern bereinigt werden, sodass später eine valide Verarbeitung und Analyse möglich wird. Entsprechend beschäftigt sich dieser erste Prozessbereich mit drei zentralen Fragestellungen:

- Welche Charakteristika möchte ich in Form von Daten und auf welche Weise erfassen?
- Wie kann ich diese so speichern, dass ich sie später weiterverarbeiten kann?
- Sind die erfassten Daten für den angestrebten Zweck geeignet?

(P2) *implementieren und optimieren* kann zu verschiedenen Zeitpunkten im Datenlebenszyklus stattfinden: Insbesondere betrifft es die Implementierung des Datenmodells in einem geeigneten Datenspeicher und die Speicherung der Daten in diesem. Andererseits ist – je nach gewählter Methode und gewähltem Werkzeug – gegebenenfalls auch bereits bei der Datenerfassung oder später bei der Analyse eine entsprechende Implementierung (und auch Optimierung), beispielsweise von einfachen Algorithmen, nötig. Entsprechend können auch bei der Optimierung verschiedene Ziele verfolgt werden, die alle anderen Bereiche des Datenlebenszyklus betreffen können. Die zentralen Fragestellungen dieses Prozessbereichs sind daher:

- Wie kann ich die Datenerfassung, -speicherung und -analyse praktisch realisieren?
- Wie kann ich das bisher Erreichte hinsichtlich konkreter Ziele verbessern?

(P3) *analysieren, visualisieren und interpretieren* befasst sich mit der Extraktion neuer Informationen aus den gewonnenen und gespeicherten Daten. Dabei wird auf verschiedene Analysemethoden und Prinzipien zurückgegriffen. Um die gewonnenen Informationen, aber auch die ursprünglichen Daten einfacher erfassbar zu machen, und zur Unterstützung der Interpretation werden häufig verschiedene Visualisierungstechniken eingesetzt. Teilweise kann auch die eigentliche Datenanalyse

mit visuellen Methoden durchgeführt werden. Der dritte Prozessbereich beschäftigt sich entsprechend mit drei zentralen Fragestellungen:

- Welche Informationen kann ich wie aus meinen Daten extrahieren?
- Wie kann ich den Menschen dabei unterstützen, das Wesentliche einfach zu erfassen?
- Welche Schlüsse kann ich aus meinen Analyseergebnissen ziehen?

(P4) *austauschen, archivieren und löschen* ist während des gesamten Erfassungs-, Speicherungs- und Analyseprozesses von Bedeutung. Dieser betrachtet das Austauschen, Archivieren und Löschen von Daten, drei Tätigkeiten, die immer relevant sind, wenn mit Daten gearbeitet wird. Es muss beispielsweise entschieden werden, wer wie Zugriff auf Originaldaten oder Analyseergebnisse bekommt, welche Daten es Wert sind erfasst und ggf. langfristig gespeichert zu werden, welche Daten anonymisiert oder pseudonymisiert werden müssen, wie Daten, beispielsweise unterstützt durch Metadaten, wieder auffindbar werden etc. Gleichzeitig markiert die Löschung von Daten jedoch auch das einzige Ende des Datenlebenszyklus, sodass es bei dieser insbesondere darauf ankommt, zu entscheiden, wann Daten gelöscht werden sollen und wie dies auf angemessene und sichere Weise geschehen kann. Damit werden in diesem Prozessbereich die folgenden Fragestellungen angegangen:

- Welche Daten möchte ich wie mit wem teilen?
- Welche Daten sollte ich langfristig archivieren?
- Wie können Daten langfristig archiviert werden?
- Wie kann ich meine Daten angemessen löschen?

In der vorgestellten Form dient das Datenkompetenzmodell primär als erste Orientierung und muss zukünftig noch weiter detailliert werden.

---

## Bedeutung von Datenkompetenz in anderen Fächern

Wie bereits dargestellt, ist die Förderung von Data Literacy zwar aufgrund fachlicher Bezüge im Interesse der Informatik, aber sie ist nicht auf diese beschränkt: Eine sinnstiftende Arbeit mit Daten findet immer in einem Kontext statt, der nicht nur aus der Informatik stammen muss, sondern oft aus anderen Fächern. Entsprechend entstehen derzeit insbesondere im Bereich der Hochschulen verstärkt Bestrebungen zur Förderung von Datenkompetenzen in interdisziplinären Bereichen. Die entsprechenden Möglichkeiten sind jedoch nicht auf Hochschulen beschränkt, sondern auch für andere Bereiche des Bildungssystems sehr spannend. Aus diesem Grund wurden erste Überlegungen und Diskussionen angestellt, die beiden vorgestellten Modelle als Basis für die interdisziplinäre Förderung von Datenkompetenzen zu verwenden. In Diskussio-

nen dazu zeigte sich, dass die Vertreter der beteiligten Fachrichtungen (u. a. Biologie, Geografie, 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 (siehe Bild 2, Seite 19) 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 eindeutig 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, da diese aus Perspektive der Informatik entstanden sind und somit ggf. Spezifika anderer Fächer natürlich ausblenden – somit ist es beispielsweise denkbar, durch weitere Inhaltsbereiche Fachspezifika abzudecken. Trotzdem ist es bereits jetzt wichtig, gemeinsam mit anderen Fächern eine interdisziplinäre Sicht auf Daten zu etablieren, indem beispielsweise Themen wie Geoinformationssysteme aus der Geografie auch aus informatischer Perspektive mitbetrachtet werden, Datenanalysen, wie sie in der Physik bei Experimenten stattfinden, auch in der Informatik als Kontext herangezogen werden oder auf Datenanalysen basierende Aspekte sprachlicher Fächer (z. B. Computerlinguistik) ebenfalls miteinbezogen werden.

## Fazit und Anwendung der Modelle im IU

In diesem Beitrag wurde – basierend auf einer kurzen Argumentation der Relevanz grundlegender Datenkompetenzen – ein umfassender Überblick über diese und das aus informatischer Perspektive zugrundeliegende Datenmanagement gegeben. Bei der Entwick-

lung der beiden vorgestellten Modelle wurde primär die Perspektive der Schulinformatik eingenommen. Erste Eindrücke eines interdisziplinären Diskurses zeigen jedoch, dass das fachlich orientierte Modell der Schlüsselkonzepte des Datenmanagements sowie das darauf basierende Datenkompetenzmodell als Basis für die anlaufende interdisziplinäre Förderung von Datenkompetenzen dienen: Aufbauend 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. Entsprechend sollen die vorgestellten Modelle helfen, eine zielgerichtete Betrachtung des Themas *Big Data* und einer gezielten Förderung der damit einhergehenden Datenkompetenzen zu betrachten, wie es in den Beiträgen „Twitterdaten analysieren – mithilfe der blockbasierten Programmiersprache SNAP!“ (Seite 54 ff.) sowie „Datenflussorientierte Big-Data-Analyse mit *Orange 3*“ (Seite 61 ff.) in diesem Heft noch praktisch skizziert wird.

Weitergehende Informationen zu den Modellen, die im Rahmen wissenschaftlicher Arbeiten (vgl. insbesondere Grillenberger, 2019a) entstanden sind, können, gemeinsam mit praktischen Beispielen zu deren Nutzung, auf der Internetpräsenz

<https://dataliteracy.education>

gefunden werden.

Dr. Andreas Grillenberger  
Freie Universität Berlin  
Didaktik der Informatik  
Königin-Luise-Straße 24–26  
14195 Berlin

E-Mail: [andreas.grillenberger@fu-berlin.de](mailto:andreas.grillenberger@fu-berlin.de)

## Literatur und Internetquellen

GI – Gesellschaft für Informatik (Hrsg.): Grundsätze und Standards für die Informatik in der Schule – Bildungsstandards Informatik für die Sekundarstufe I. Erarbeitet vom Arbeitskreis „Bildungsstandards“ unter Federführung von Hermann Puhlmann – Empfehlungen der Gesellschaft für Informatik e. V. vom 24.01.2008. In: LOG IN, 28. Jg. (2008), Nr. 150/151, Beilage.  
<https://t1p.de/csqx>

Brewer, E.: CAP Twelve Years Later – How the “Rules” Have Changed. 2012. In: *Computer*, 45. Jg. (2012), Heft 2, S.23–29.  
<https://t1p.de/k980>

Grillenberger: Von Datenmanagement zu Data Literacy – Informatikdidaktische Aufarbeitung des Gegenstandsbereichs Daten für den allgemeinbildenden Schulunterricht. Berlin: Freie Universität Berlin, Fachbereich Mathematik und Informatik, 2019a (Dissertation).  
<http://dx.doi.org/10.17169/refubium-1932>

Grillenberger, A.: Detaillierte Beschreibung der Schlüsselkonzepte des Datenmanagements. Auszug aus A. Grillenberger: Von Datenmanagement zu Data Literacy – Informatikdidaktische Aufarbeitung des Gegenstandsbereichs Daten für den allgemeinbildenden Schulunterricht.

Berlin: Freie Universität Berlin, Fachbereich Mathematik und Informatik, 2019b (Dissertation), S.257–283.  
<https://t1p.de/pva8>

Grillenberger, A.; Romeike, R.: Datenmanagement als Thema für den Informatikunterricht. In: LOG IN, 37. Jg. (2017), Heft 187/188, S.44–52.

Mayer-Schönberger, V.; Cukier, K.: Big Data – Die Revolution, die unser Leben verändern wird. München: Redline Verlag, 2013.

Ridsdale, Ch.; Rothwell, J. u. a.: Strategies and Best Practices for Data Literacy Education – Knowledge Synthesis Report. Halifax (CAN): Dalhousie University. 2015.  
<https://t1p.de/q3no>

Scholtes, I.: Data Science in den Sozialwissenschaften. In: GI – Gesellschaft für Informatik (Hrsg.): Data Literacy und Data Science Education – Digitale Kompetenzen in der Hochschulausbildung. Berlin: GI Berlin, 2018, S.40–43.  
<https://t1p.de/1kix>

Alle Internetquellen wurden zuletzt am 18. Februar 2021 geprüft und können auch aus dem Service-Bereich des LOG IN Verlags (<https://www.log-in-verlag.de>) heruntergeladen werden.

Anzeige

„Ich träume davon, **zur Schule gehen** zu können.“

**kinder not hilfe**

60 Jahre  
Gemeinsam wirken

[kindernothilfe.de/patenschaft](http://kindernothilfe.de/patenschaft)

DZI  
Spenden-Siegel