

Hagenhoff, Svenja

Research Report

Gegen die Diskussion mit den drei Unbekannten Daten, Algorithmen und Digitalisierung

Erlanger Beiträge zur Medienwirtschaft, No. 13/2020

Provided in Cooperation with:

Friedrich-Alexander University of Erlangen-Nuernberg (FAU), Institute for the Study of the Book, Professorship of E-Publishing and Digital Markets

Suggested Citation: Hagenhoff, Svenja (2020) : Gegen die Diskussion mit den drei Unbekannten Daten, Algorithmen und Digitalisierung, Erlanger Beiträge zur Medienwirtschaft, No. 13/2020, Friedrich-Alexander-Universität Erlangen-Nürnberg, Institut für Buchwissenschaft, Professur für Buchwissenschaft, insb. E-Publishing und Digitale Märkte, Erlangen, <https://nbn-resolving.de/urn:nbn:de:bvb:29-opus4-150014>

This Version is available at:

<https://hdl.handle.net/10419/227581>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/3.0/de/>

Beitrag Nr. 13/2020

Hrsg.: Svenja Hagenhoff

Svenja Hagenhoff

Gegen die Diskussion
mit den drei Unbekannten
Daten, Algorithmen und Digitalisierung

Erlanger Beiträge zur Medienwirtschaft

Herausgegeben von Svenja Hagenhoff
Friedrich-Alexander-Universität Erlangen-Nürnberg
Institut für Buchwissenschaft
Katholischer Kirchenplatz 9
91054 Erlangen

Erlangen Contributions to Media Management and Media Economics

Edited by Svenja Hagenhoff
Friedrich-Alexander University of Erlangen-Nürnberg
Institute for the Study of the Book
Katholischer Kirchenplatz 9
91054 Erlangen / Germany



Tel. +49 (0) 9131 / 85-24700
Fax +49 (0) 9131 / 85-24727
<https://buchwissenschaft.phil.fau.de/>



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0
Germany License

Abstract

The inflationary use of currently prominent terms, such as data centration or the digital age, without the effort of naming and explaining how the phenomena they refer to are to be understood and by which characteristics they are characterized, does nothing more than obscure meaning without any chance of gaining insight or well-founded opinion. Algorithms, data, digitization, and are three of the candidates that are frequently used but only rarely explained in a well-founded and differentiated manner. This article comes into this desideratum. It deals with these 'three unknowns' and attempts to grasp and substantiate the terms and the phenomena they refer to and to formulate what exactly is 'new' and what constitutes the quality of difference to previous states (age without digitization).

Inhaltsverzeichnis

Abbildungsverzeichnis.....	III
Abkürzungsverzeichnis.....	IV
1 Motivation und Positionierung des Beitrags.....	1
2 Algorithmus	2
2.1 Begriff und Konzeptualisierung.....	2
2.2 Eigenschaften von Algorithmen.....	2
3 Daten.....	4
3.1 Begriffe und Konzeptualisierung.....	4
3.2 Arten und Eigenschaften von Daten	6
3.3 Datenzentrierung, Datengetriebenheit oder Daten als Basis von Etwas	9
4 Digitalisierung.....	11
4.1 Begriff und Konzeptualisierung.....	11
4.2 Ausprägungen von Digitalisierung 1: Wertschöpfung	12
4.3 Ausprägungen von Digitalisierung 2: Relation zwischen Entitäten und Bits	14
5 Versuch einer Synopse und Identifikation von Synapsen.....	17
5.1 Was ist jetzt eigentlich neu?	17
5.2 Und nun?	18
Literaturverzeichnis.....	19

Abbildungsverzeichnis

Abbildung 1: EVA-Prinzip und Zusammenhang von Algorithmus-Qualität und Daten-Qualität	4
Abbildung 2: Arten und Eigenschaften von Daten.....	7
Abbildung 3: Elektronischer Datenaustausch zwischen Verlag und Bibliothek.....	8
Abbildung 4: Ausprägungen von Digitalisierung in der Wertschöpfung.....	14
Abbildung 5: Beschreibung einer Bibliothek.....	15
Abbildung 6: Asset Management.....	16
Abbildung 7: Rekonstruktionen (Drahtgittermodell links und texturiertes Modell rechts).....	16

Abkürzungsverzeichnis

EVA Eingabe, Verarbeitung, Ausgabe

GWB Gesetz gegen die Beschränkungen des Wettbewerbs

1 Motivation und Positionierung des Beitrags

Anfang September 2020 hat das Bundeswirtschaftsministerium den Entwurf für das ›Digitalisierungsgesetz‹ als Ergänzung zum Gesetz gegen die Beschränkungen des Wettbewerbs vorgelegt (GWB-Digitalisierungsgesetz). Begründet wird die Notwendigkeit der Ergänzung damit, dass eine Wettbewerbspolitik für das digitale Zeitalter adäquate Regeln und Instrumente brauche. Insbesondere *Daten* hätten eine immer stärkere Bedeutung als Wertschöpfungsfaktor (Bundesministerium für Wirtschaft und Energie 2020). In einem Kommentar hierzu stellt Gerpott (2020) fest, dass weder der Gesetzesentwurf noch dessen Begründung Auskunft über die Weite der Auslegung des Begriffs Daten gäben. Ein Blick in den Entwurf zeigt, dass nicht nur die Weite des als zentral erachteten Phänomens nicht erläutert wird, sondern dass überhaupt nicht ausgeführt wird, was unter Daten konkret verstanden werden könnte. Gerpott (2020) fügt hinzu, dass ohnehin nicht die Daten, sondern die Algorithmen der eigentliche Wettbewerbsfaktor seien und sich daher das Gesetz nicht auf die Regulierung des Zugriffs von Wettbewerbern auf Daten beschränken sollte, sondern ein präzise konturiertes Recht auf den Zugang zu Programmen umfassen solle. Hofstetter (2018, Klappentext) verdichtet dieses in der Formulierung »Big Data speichert unser Verhalten, künstliche Intelligenzen analysieren unsere Absichten«.

Damit sind in wenig Text – Kommentar und Ankündigung eines Gesetzesentwurfs – drei Begriffe verwendet – Daten, Algorithmen, Digitalisierung – die aktuell geradezu inflationär verwendet werden (auch: Datenzentrierung, Datafizierung oder Algorithmisierung), jedoch ohne die Mühe des Benennens und Erläuterns, wie die damit bezeichneten Phänomene zu fassen seien und durch welche Eigenschaften sie charakterisiert wären. Dieses leistet jedoch nichts anderes als Bedeutungsverunklarung¹ ohne eine Chance auf Gewinn an Erkenntnis oder wohlbegründeter Meinung.

Dieser Beitrag stößt in dieses Desiderat. Er behandelt die ›Drei Unbekannten‹ Daten, Algorithmen und Digitalisierung und versucht sich daran, die Begriffe bzw. die damit bezeichneten Phänomene zu fassen und zu fundieren (Kapitel 2 bis 4). Kapitel 5 liefert den Versuch einer Synopse mit Fragen danach, was genau nun ›neu‹ ist (Zeitalter der Digitalisierung) im Vergleich zu vorangegangenen Zuständen (Zeitalter ohne Digitalisierung) und zeigt die Notwendigkeit interdisziplinärer Forschung aus der Perspektive der Medien- und Kommunikationsforschung Forschung auf.

¹ Begriff nach Loleit (2004, S. 207).

2 Algorithmus

2.1 Begriff und Konzeptualisierung

Verdichtet auf das, was verschiedenen Definitionsversuchen (z. B. Boockmeyer et al. 2017; Drösser 2016, Pos. 223; Gillespie 2014, S. 1; Garey und Johnson 1979, S. 4; Kowalski 1979, S. 424) von Algorithmus gemein ist, lässt sich dieser definieren als eine Handlungsanweisung. Sie kann unabhängig von Software oder Computern (in natürlicher Sprache) formuliert und auch ausgeführt werden, wird üblicherweise aber stark formalisiert notiert. Die einzelnen Schritte einer Anweisung werden in Reihenfolgen ausgeführt, ob ein spezifizierter Schritt ausgeführt wird oder nicht kann von zahlreichen Bedingungen abhängig gemacht werden, und einzelne Schritte müssen nicht zwingend hintereinander, sondern können auch parallel ablaufen. Einzelne Schritte können mehrmals hintereinander durchlaufen werden, solange, bis eine Bedingung eingetreten ist. Damit Algorithmen (semi-)automatisch ausgeführt werden können ist es erforderlich, sie in Software zu überführen, sie also maschinenlesbar zu machen.

Algorithmen sind Artefakte. Sie fallen nicht aus dem Himmel, sondern sind menschengemacht und arbeiten daher auch nicht nach ihren »eigenen Kriterien« (dieses bei Schmidt (2013, S. 53). Algorithmen sind daher auch nicht neutral.² Auch selbstlernende Algorithmen benötigen eine Trainings- und eine Testphase, in denen Menschen verschiedene Aufgaben übernehmen, wie z. B. das Säubern und Transformieren von Daten oder die Wahl grundsätzlich anzuwendender Modelle.

2.2 Eigenschaften von Algorithmen

Ein Algorithmus ist anwendbar auf alle Probleme der gleichen Problemklasse (*Allgemeingültigkeit*). Er ist abstrakt und nicht konkret. Ein Sortieralgorithmus muss drei Listeneinträge genauso sortieren wie 30.000. Zwar können für die gleiche Problemklasse unterschiedliche Algorithmen vorliegen, dennoch müssen alle Verfahren das Problem prinzipiell lösen. Ein Algorithmus ist von endlicher Länge (*Fintheit*), da er ansonsten nicht abgearbeitet werden kann. Er besteht daher zwingend aus einer endlichen Menge an Schritten, die in endlicher Zeit durchlaufen werden. Nicht zu verwechseln mit der Endlichkeit der Beschreibung des Algorithmus ist die Endlichkeit des Prozesses (*Terminiertheit*). Terminierende Algorithmen kommen von selber zu einem Ende. Das Ende eines Prozesses im Sinne von *fertig* muss definiert sein, so dass das Erreichen des Endes feststellbar ist, wie z. B. bei Brettspielen der Fall. Nicht-terminierende Algorithmen werden für Vorgänge benötigt, die nicht abbrechen sollen, wie z. B. die Steuerung eines Hochofens im Eisenwerk. Wenn ein Algorithmus bei gleichen Voraussetzungen das gleiche Resultat liefert ist es ein *determinierter* Algorithmus. Für etliche Problemklassen ist das eine zwingende Forderung. Drei plus vier ist immer sieben und nicht nur manchmal. Nicht-determinierte Algorithmen liefern aus einer Menge möglicher Ergebnisse irgendeines, sie treffen also eine Auswahl, wie z. B. ein Empfehlungsmechanismus. Determiniertheit ist nicht gleichzusetzen mit *Determinismus*: letzteres bedeutet, dass auf dem Weg zum Ergebnis jeder einzelne Schritt

² Zur Kritik am Neutralitätsmythos vgl. Mowshowitz (1984) oder Weizenbaum (1977).

für alle Instanzen eines Problems durch Vorschriften erschöpfend bestimmt ist. Ein solcher Algorithmus folgt immer exakt dem gleichen, vorgegebenen Weg von der Eingabe zur Ausgabe, alle Zwischenschritte nehmen nach der gleichen Eingabe immer die gleichen Zustände an.³ Nicht deterministische Algorithmen folgen auf dem Weg zur Lösung unterschiedlichen Wegen. Die jeweils aktuelle Konfiguration des Algorithmus kann nicht bestimmt werden, da es keinen festen funktionalen Zusammenhang zwischen Inputvariablen und Zielvariable gibt (Pietsch und Wernecke 2017), die Routenplanung ist eine solche Problemklasse. Nicht-deterministische Algorithmen werden vor allem benötigt für Probleme von irreduzibler Komplexität. Hierfür können selbstlernende Algorithmen eingesetzt werden, deren Arbeitsweise nah an der Art und Weise ist, wie der Mensch seine unstrukturierten Probleme löst: auch die Routinen der menschlichen Denkleistung sind nicht einsehbar, häufig nicht präzise und widerspruchsfrei, oftmals gar nicht explizierbar und damit nicht transparent (Pólya 2010).

An Algorithmen werden die Anforderungen der Effizienz und der Korrektheit gestellt. Effizient sind solche Algorithmen, die für die Lösung gleicher Probleme weniger Laufzeit und weniger Speicherplatz benötigen. Unter unten stehendem Link⁴ kann man verschiedenen Sortieralgorithmen bei der Arbeit zusehen: Man sieht, dass unterschiedliche Verfahren in Abhängigkeit von der Ausgangssituation sowie der Mächtigkeit des Sortierproblems unterschiedlich viel Zeit benötigen.

Ein Algorithmus soll korrekt arbeiten. Das bedeutet, dass er die gestellte Aufgabe richtig löst. Die Herausforderung besteht darin, dass die Korrektheit faktisch häufig gar nicht geprüft werden kann: Auf empirischem Wege durch Testen lässt sich nur die Anwesenheit, aber nicht die grundsätzliche Abwesenheit von Fehlern aufzeigen (Dijkstra 1972). Formal-theoretische Korrektheitsbeweise wiederum erfordern, dass das zu lösende Problem vollständig und präzise spezifiziert ist (Ottmann und Widmayer 2017, S. 2), was bei Problemklassen mit irreduzibler Komplexität gar nicht gelingt. Zudem existieren Problemklassen, für die *korrekt* gar nicht objektiv definierbar ist (Schöning 2008, S. 117). Insofern ist (Schmidt 2013, S. 73) zu widersprechen, der in Bezug auf Social Bots formuliert, dass deren Einsatz keine grundlegenden Bedenken aufwirft, »sofern sie technisch sauber programmiert sind, also keinen fehlerhaften Output produzieren.« Die grundsätzliche Klassifikation eines Outputs als »Fehler« setzt ein allgemeingültiges Verständnis davon voraus, was als zulässige Lösung gilt, das ist aber nicht in allen Anwendungsfällen eindeutig bestimmbar, wie das Beispiel mit der Auswahl von Büchern zeigt: Warum der Buchhändler dieses Buch empfiehlt, jenes aber nicht, ist nicht nur schwer formulierbar, das Ergebnis ist stochastischer Art. Der Algorithmus muss »nur« das Problem lösen, überhaupt etwas aus einer sehr großen, meist unüberschaubaren Menge an Objekten auszuwählen, also in diesem Sinne das *Richtige* zu tun. Welches die adäquaten Parameter sind (möglich auch: Zufall) ist zeitabhängiger Gegenstand von Vereinbarung auf den ver-

³ Zu gültigen und determinierten Zuständen bei der Abwicklung des Ausleihprozesses in einer Bibliothek vgl. Hagenhoff (2015-2020, Abb. 2).

⁴ <https://www.toptal.com/developers/sorting-algorithms/> [26.11.2020].

schiedenen Ebenen sozialer Systeme. Pietsch und Wernecke (2017, S. 20) sprechen diesbezüglich von *phänomenologischer Berechenbarkeit*, der ein anderer Wahrheitsbegriff zugrundliegt als bei einer epistemischen Berechenbarkeit.⁵

Da Algorithmen und Daten zwingend zusammengehören⁶, sind die Ergebnisse eines Algorithmus zwangsläufig abhängig vom Input: Sogar eindeutig korrekte Algorithmen liefern falsche Ergebnisse, wenn die verarbeiteten Daten von schlechter Qualität sind. Dieser Zusammenhang, der sich in dem simpel anmutenden EVA-Prinzip niederschlägt und in der *Garbage in – Garbage out*-Phrase verdichtet, wird in der Algorithmen Diskussion häufig übersehen. Zudem produzieren auch deterministische Algorithmen divergente Ergebnisse, wenn sich die Datenbasis im Zeitablauf ändert, was bei historisierender Datenhaltung der gewünschte Fall ist.

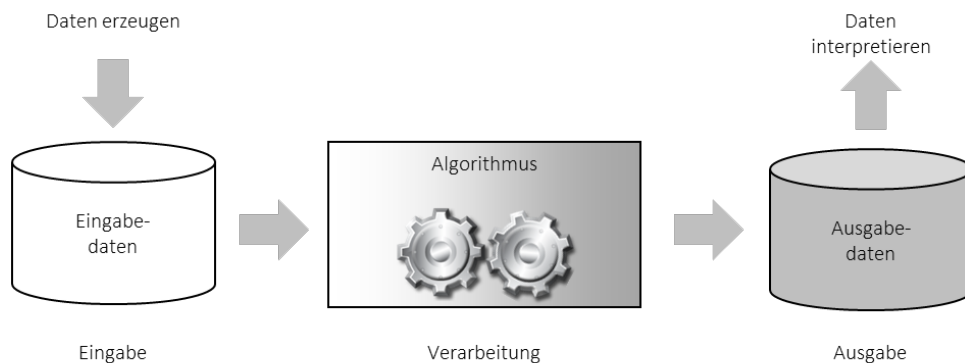


Abbildung 1: EVA-Prinzip und Zusammenhang von Algorithmus-Qualität und Daten-Qualität

3 Daten

3.1 Begriffe und Konzeptualisierung

Etymologisch bedeutet das Wort Datum als Singular zu Daten »gegeben« oder »das, was gegeben ist« (Pfeifer 1993a, zur Historiographie der Verwendung des Begriffs Rosenberg 2013). Daten werden in dieser Interpretation verstanden als objektive Tatsachen, Messwerte, die Gegebenheiten der Realwelt beschreiben (Kitchin 2014, S. 2). Rosenberg (2013) differenziert etymologisch die Begriffe *fact* (*facere*: to do) und *evidence* (*vidēre*: to see) und führt aus: »facts are ontological, evidence is epistemological, data is rhetorical. A datum may also be a fact, just as a fact may be evidence. But, from its first vernacular formulation, the existence of a datum has been independent of any consideration of corresponding ontological truth. When a fact is proven false, it ceases to be a fact. False data is data nonetheless«. In etlichen Disziplinen hat sich allerdings eine Konzeptualisierung von

⁵ Zur Veränderung des Logik-Begriffs im Sinne von wahr/falsch oder ja/nein sowie zur Unlogik des menschlichen Denkapparats Tschudi und D’Onofrio (2018), zur Art des menschlichen Schließens auf Basis verschiedener kommunikationstauglicher Zeichentypen Keller (1995), knapp bei Reischer (2006, S. 54ff.); zur Eindeutigkeit von Frege’schen Begriffen und solchen mit unscharfen Rändern ebenfalls Keller (1995, S. 95); Reischer (2006, S. 59).

⁶ Hier ist dann Gerpott (2020) zuzustimmen bei seiner Feststellung, dass der Gesetzesentwurf eine zentrale Komponente des digitalen Zeitalters, nämlich die Algorithmen, nicht thematisiert.

Daten durchgesetzt, die nur in Relation zu anderen Konzepten und Begriffen gefasst werden kann, dieses sind Zeichen, Information, Wissen und manchmal Weisheit sowie Inhalt. So werden bspw. in der Wirtschaftsinformatik, der Medieninformatik, der BWL und in der Informations- und Bibliothekswissenschaft mehr oder weniger komplexe Abgrenzungsketten erstellt, in denen durch Veredelungsbeigaben (Reichert 2018, S. 21), wie Syntax, Semantik, Kontextualisierung oder Vernetzung, ein Konzept in ein höherwertiges anderes transformiert wird. Am wenigsten differenziert geht die Informatik vor, die unter Daten maschinell verarbeitbare, verschiedenartig formalisierte Repräsentationen von Informationen versteht. Informationen wiederum sind Angaben zu realweltlichen Sachverhalten, Objekten, Vorgängen, Fakten und Gedanken, sie stellen eine Abstraktion von Daten dar (z. B. Gumm und Sommer 2008, S. 4). Trotz der Unterschiede im Detail ist allen Stufenmodellen gemein, dass auf höheren Ebenen kognitive Handlungsmuster entstehen (Laudon et al. 2016, S. 641), wobei Daten in allen Modellen auf einer unteren Stufe angesiedelt sind. Gleichwohl ist eine objektive Grenzziehung zwischen den Stufen der verschiedenen Modelle rein idealtypischer Art. Die Grenzen sind intersubjektiv instabil, da der Zustandswechsel von einer Stufe auf die andere vom jeweils interpretierenden System abhängt (Rieder und Röhle 2012; Keller 1995, S. 117).

Ballsun-Stanton (2010) hat auf empirischem Wege mittels Interviews drei Konzeptualisierungen für Daten erarbeitet. *Data as hard numbers* sind wahre Aussagen über die Realität, die objektiv prüfbar sind. Mit mehr und differenzierten Daten lässt sich in dieser Denkweise die Realität immer genauer beschreiben und immer genauer bestimmen (Reichert 2018, S. 19). Vernachlässigt wird in solchen Konzepten, dass Aussagen über die Realität immer eines kulturell vereinbarten Zeichensystems als Code bedürfen, selbst die Zahl Null ist nicht frei von Kontext nutz- und interpretierbar (hierzu Posner 1983). Ebenso sind die Ergebnisse von Messungen immer Produkte technischer Infrastrukturen und experimenteller Anordnungen (Reichert 2018, S. 22), so dass die Erhebung der Daten nicht bedingungslos stattfindet. *Data as recorded observations* stellen subjektive Wahrnehmungen dar, die zur Interpretation Kontexte benötigen. Dieser Ansatz anerkennt, dass Daten nur in komplexen Begründungs- und Argumentationsverfahren (Reichert 2018, S. 21) entstehen und vor dem Hintergrund dieser Bedingungen nutzbar sind. *Data as bits* dienen der computerbasierten Kommunikation. Während Daten sich in den beiden ersten Konzepten immer auf reale Phänomene beziehen, steht bei letzterem der semiotische Charakter im Vordergrund (Voß 2013b). Abstrakter könnte das Konzept auch als *Data as communication* formuliert werden. Daten würden dann unterschiedliche, menschenlesbare Formen annehmen (z. B. Buchstaben), die lediglich zum Zwecke des maschinellen Transports auf Bitfolgen reduziert werden.

In den Geistes- und Kulturwissenschaften spielt der Begriff bis dahin keine prominente Rolle. Auch im Zuge des Aufkommens der Digital Humanities (als Standardwerk z. B. Jannidis et al. 2017) wird der Begriff Daten selten erläutert, eine Ausnahme ist Owens (2011), der sich aber sehr unscharf ausdrückt. Als Differenzbeitrag zu dem bereits dargebotenen soll hier seine Konzeptualisierung von *Daten als interpretierbare Texte* erwähnt werden. »Data is created for an audience. Humanists can, and should interpret data as an authored work and the intentions of the author are worth consideration and exploration«. In dieser Logik sind Daten das Material, aus dem Kommunikate bestehen, diese entspricht einer der Sichten bei Hagenhoff (2015-2020). Eine gewisse Aufmerksamkeit hat noch

Floridi (2005, 2009, 2010) mit seinen Arbeiten zum Begriff und Konzept Information erlangt. Er definiert Daten in Anlehnung an das Konzept der Entropie der Informationstheorie nach Shannon als etwas, was einen ›Unterschied‹ repräsentiert: »datum =def. x being distinct from y, where x and y are two uninterpreted variables and the relation of 'being distinct', as well as the domain, are left open to further interpretation.« (Floridi 2010, S. 22).

Standardwerke der Kommunikations- und Medienwissenschaft enthalten den Begriff *Daten* (meistens noch) nicht. Das *Lexikon Kommunikations- und Medienwissenschaft* (Bentele et al. 2013) enthält etliche Komposita mit Daten*, den Begriff Daten selber aber nicht; gleiches gilt für das *Gabler Lexikon Medienwirtschaft* (Sjurts 2011) sowie für *Reclams Sachlexikon des Buches* (Rautenberg 2015). Der Begriff befindet sich aber in Kombination mit *Big* (Big Data) bzw. *Open* (Open Data) in Heesen (2016), allerdings auch hier, ohne dass Daten als Begriffskern dieser größeren Konzepte erläutert würde. Im *Handbuch Medienökonomie* ist ein Aufsatz mit Begriffserläuterung enthalten (Hagenhoff 2015-2020).

3.2 Arten und Eigenschaften von Daten

Um eine weitere Vorstellung davon zu erhalten, was Daten sein könnten, ist es sinnvoll, die in der Literatur diskutierten Eigenschaften von Daten näher anzuschauen. In dem nachstehenden morphologischen Kasten sind die Klassifizierungsmöglichkeiten von Daten zusammengestellt. Die (ausführlichen) Beschreibungen hierzu finden sich bei Fahsel et al. (2017); Palmetshofer et al. (2017); Otto et al. (2016); Hagenhoff (2015-2020); Kitchin (2014); Müller und Lenz (2013); Brosche (2014); Bensberg und Weiß (1999).⁷

⁷ In den genannten Quellen werden Steuerdaten und Passivdaten nicht besprochen: Steuerdaten veranlassen Aktivitäten oder lenken Datenströme, Passivdaten sind diejenigen, die verarbeitet oder transportiert werden.

Kriterium	Ausprägung			
<i>Formbezogene Kriterien</i>				
Erscheinungsform (Teil 1)	Statisch: Schriftlich	Statisch: bildlich	Dynamisch: bildlich	Dynamisch: akustisch
Zeichentyp (Teil 1)	Symbolisch		Ikonographisch	
Zeichentyp (Teil 2)	Alphabetisch	Numerisch	Sonderzeichen	Alphanumerisch
Zeichentyp (Teil 3)	Numerisch (quantitativ)		Nicht-numerisch (qualitativ)	
Formatierung	Formatiert		Nicht formatiert	
<i>Ursprungsbezogene Kriterien</i>				
Zielorientierung der Erhebung	Gezielt erhobene Daten (captured data)		Daten als ‚Beifang‘ (exhausted data)	
Erzeugungsform	Daten als Ergebnis von Erhebung		Daten als Ergebnis von Verarbeitung	
Erhebungsmöglichkeiten	1st-Party	2nd-Party	3rd-Party	
Nutzeraktivität	Transaktionen		Interaktionen	
<i>Verarbeitungs- und nutzungsbezogene Kriterien</i>				
Stellung im Verarbeitungsprozess	Eingabedaten		Ausgabedaten	
Basisoperationen	Rechendaten	Ordnungsdaten	Identifikationsdaten	
Aktivitätsgrad	Steuerdaten (veranlassen Aktivitäten)		Passivdaten (werden verarbeitet)	
Verwendungszweck (Teil 1)	Zustandsdaten: Stammdaten	Zustandsdaten: Bestandsdaten	Abwicklungsdaten: Vormerkdaten	Abwicklungsdaten: Bewegungsdaten
Verwendungszweck (Teil 2)	Operative Daten		Dispositive Daten	
Funktion (Teil 1)	Metadaten		Nutzdaten	
Funktion (Teil 2)	Identifikationsdaten	Eigenschaftsdaten	Metadaten	
Grad der Strukturierung	Strukturiert		Unstrukturiert	
Grad der Beständigkeit	Transient		Persistent	
Zeitbezug	Temporal / historisierend		Nicht-temporal	
<i>Governance-bezogene Kriterien</i>				
Zugänglichkeit	Frei zugänglich		Nicht frei zugänglich	
Datenherkunft	Privat	Öffentlich	Gemeinschaftlich	
Datenerzeugung	Freiwillige Angaben	Aus Beobachtung	Abgeleitet	

Abbildung 2: Arten und Eigenschaften von Daten

Wie gut Daten (automatisiert) verarbeitet werden können hängt von ihrer objektiven Beschaffenheit ab sowie auch von ihrer Qualität⁸ in Bezug auf ein angestrebtes Ziel; ein Aspekt, der in Diskussionen in der Regel völlig vernachlässigt wird (Meyer-Wegener 2017).

Von den verschiedenen »V«, die das Phänomen *Big Data* je nach Quelle beschreiben (sollen), werden hier drei besprochen, um die operativen Herausforderungen im Umgang mit Daten zu verdeutlichen: Daten sind *vielfältig* (Variety) hinsichtlich zweier Aspekte (Meyer-Wegener 2017, S. 102): zum einen sind sie in Bezug auf Erscheinungsform, Zeichentyp und Strukturiertheit inhomogen, daran ändert auch die »einheitliche« Binärcodifizierung nichts. Relativ einfach automatisiert zu handhaben sind Daten, die strukturiert z. B. in Tabellenform oder XML vorliegen. Die Bedeutung der Daten wird in den Spaltenköpfen der Tabelle oder in den XML-Tags benannt. Dennoch bestehen erhebliche Herausforderungen einfach nur deswegen, weil gleiche Sachverhalte divergent abgelegt werden. Das Kalenderdatum kann z. B. als *15.03.2016* oder *03/15/16* oder als *Luise* repräsentiert sein (Meyer-Wegener 2017, S. 112); für eine automatisierte Routine sind das drei verschiedene Zeichenketten, die in verschiedenen Bytefolgen codifiziert sind. Komplexer zu verarbeiten sind semistrukturierte Daten in Textform. Texte enthalten zumindest noch größere Strukturelemente wie Überschrift, Inhaltsverzeichnis oder Abstract, und sie sollten den Regeln der Grammatik gehorchen, dennoch sind sie strukturell von

⁸ Hierzu differenziert Fahsel et al. (2017) auf der Basis von Glaß und Schmidt (2011); Rohweder et al. (2008); Wang und Strong (1996).

vielen Freiheitsgraden gekennzeichnet. Die größte Herausforderung stellen völlig unstrukturiert vorliegende Daten (oder ggf. besser: Zeichen) dar, wie sie sich als Ansammlung von Pixeln in Bildern manifestieren. Bilder weisen keine Syntax auf (Scholz 2004, S. 125 zitiert nach Krämer 2005) im Sinne einer abstrakt formulierbaren Regelmäßigkeit und Konstruktionsanweisung.

Zum anderen sind Daten semantisch höchst divergent. Daten repräsentieren in unterschiedlichen (Teil-)Domänen (Kulturen, Epochen, ...) unterschiedliche realweltliche Konzepte, die in divergenter Differenziertheit erfasst und unterschiedlich bezeichnet werden: Nachstehend wird ein Publikationsdatensatz, der z. B. einen Buchtitel beschreibt, vom Anwendungssystem eines Verlags zu den Anwendungssystemen der Bibliotheken übertragen, eine auf den ersten Blick triviale und auch völlig langweilig anmutende, dennoch fordernde Aufgabe. Beide Domänen nutzen unterschiedliche semantische und auch syntaktische Standards, um den Entitytyp *Buchtitel* zu beschreiben. Um die Datensätze vollständig automatisiert zu transferieren und nutzbar zu machen wird zunächst ein Mapping benötigt, das die verschiedenen Standards konzeptionell aufeinander abbildet, danach kann eine Routine die Transformation der einzelnen Datensätze im Tagesgeschäft durchführen (hierzu grundsätzlich Schumann-Giesler 1997). Hinzu kommt die Definition eines Aktions-Reaktions-Regelsets, das die Ursache des erforderlichen Datenaustauschs und seine Konsequenzen im Zielsystem festlegt.

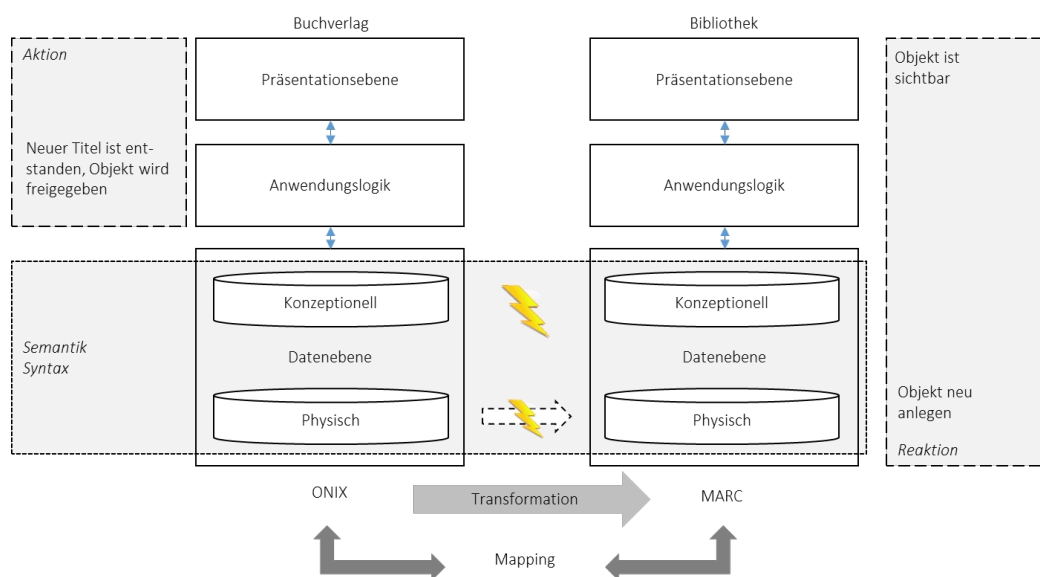


Abbildung 3: Elektronischer Datenaustausch zwischen Verlag und Bibliothek

Daten müssen glaubwürdig, gültig oder wahrhaftig sein (Veracity oder Validity). Großen Einfluss hierauf hat die Quelle der Daten und ihr Entstehungskontext: Daten können bewusst oder unbewusst unvollständig oder fehlerhaft erfasst worden sein. Unzulänglichkeiten müssen gleich bei der Datenerfassung erkannt werden, da sich ansonsten Fehler durch die gesamte Kette der Datenverarbeitung weiter hindurchziehen. Von einem Function Creep spricht man, wenn Daten für Zwecke verarbeitet werden, für die sie nicht erhoben wurden (Camenisch et al. 2015, S. 140). Wenn *Sinn* oder *Verstehen* nur durch Kontextualisierung, Interpretation und Kognition entstehen kann (hierzu mit Bezug zu Kommunikation Schmidt 1994, S. 615; Reischer 2006, Kap. 2.3), erzeugt das zweckfreie

Sammeln (»Erstmal einfach alles speichern«, Meyer-Wegener 2017) großer Datenmengen ggf. viel Rauschen aber wenig *Nutzen* oder *Wert* (Value).

3.3 Datenzentrierung, Datengetriebenheit oder Daten als Basis von Etwas

Daten werden eine herausragende Rolle als wertschöpfende Ressource zugewiesen: »Daten sind das neue Öl: Im Zeitalter der Digitalisierung besitzen sie einen immer größeren Anteil am Wertschöpfungsprozess von Produkten und Dienstleistungen« (Möller et al. 2017, S. 0).

Otto et al. (2016, S. 10) zeichnen eine Evolutionskette der Bedeutung von Daten für Unternehmen nach. In den Anfängen der (elektronischen) Datenverarbeitung waren Daten als Abbilder von Zuständen das *Ergebnis von Prozessen*. Bis in die in die 1980er Jahre hinein war die Situation dadurch gekennzeichnet, dass der Informationsfluss der Realität (z. B. dem Materialfluss) stets hinterherlief, so dass eine erhebliche zeitliche Lücke zwischen Realzuständen und ihren Abbildern in informationsverarbeitenden Systemen bestand. Beniger (1986) hat diesen Umstand für das 19. Jahrhundert differenziert beschrieben und als Control Crisis bezeichnet. Seiner Argumentation nach sind bereits damals zahlreiche technische und organisatorische Innovationen zur Bereitstellung und schnelleren Distribution von Informationen vor allem in den Bereichen Massenpresse und Telekommunikation entstanden um diese Kreise zu entschärfen, und dieses ist nach Argumentation Benigers die Basis für die Informationsgesellschaft des 20. Jahrhunderts (hierzu auch Carr 2008). Seit dem Aufkommen von ERP-Systemen in den 1980er Jahren ist eine Verarbeitung von Daten in Echtzeit möglich, so dass die Datenverarbeitung und die hierdurch gesteuerten Prozesse einen höheren Integrationsgrad erreichen (*Daten als Befähiger von Prozessen*). Seit ungefähr dem Jahrtausendwechsel gelten *Daten als Befähiger der Produkte*: Die Leistungsangebote von Unternehmen werden stärker individualisiert und auf die Bedarfe von Kunden zugeschnitten. Hierfür sind differenzierte Daten über Kunden und deren Produktnutzung vonnöten. In der letzten Stufe der Entwicklung bestehen die problemlösenden Leistungsangebote von Unternehmen selber aus Daten (*Daten als Produkt*). Daten und ihre Verarbeitung dienen nicht mehr alleine dazu, das unternehmerische Geschehen abzubilden und hierüber Prozesse zu steuern, sondern sie sind das Material aus dem der Problemlöser besteht. Mit diesem Szenario wird der Begriff der Datenzentrierung verbunden. Nach den Recherchen von Dorfer (2016) wird dieser erstmalig von Tempich und Rieger (2007) verwendet. Er bezeichnet dort im engen Sinne ein Geschäftsmodell, bei dem aus Datenbeständen immaterielle Werte geschöpft werden. Diese rein immateriellen Werte sind definiert als zweckdienliche Informationen, die die alleinige Basis für Monetarisierungen darstellen. Die Autoren bezeichnen die Akteure nach Hagel und Rayport (1999) als Infomediaries und das Ergebnis einer solchen Wertschöpfung als Pure Play Data Product. Auch Schulz und Dankert (2016) behandeln in ihrer Studie den Akteur Informationsintermediär, und verstehen darunter alle Akteure, die einen relevanten Zugangspunkt für die Erschließung von Informationen anbieten, in dem sie digital vorliegende Inhalte bündeln oder ordnen. Bei den Autoren gehören Produzenten redaktionell erstellter Inhalte (»Medienunternehmen«) gleichermaßen wie bei Tempich und Rieger nicht in diesen Typ von Geschäftsmodell, da es in Bezug auf die Vielfalt der wahrnehmbaren Inhalte weniger um die Frage ginge, was verbreitet wird, sondern

darum, was gefunden und verarbeitet werden kann, diese Aufgabe erfüllen datenzentrierte Geschäftsmodelle. In dieser Sicht geht es bei diesen Geschäftsmodellen rein um das Erzeugen von Markttransparenz in arbeitsteiligen Ökonomien mit dezentralen Organisations- und Koordinationsmechanismen, wie es z. B. auch klassische Makler seit Jahrhunderten leisten (hierzu z. B. Krajewski 2012). Diese Sicht ist identisch mit dem Konzept von Plattformökonomien (hierzu z. B. Bundesministerium für Wirtschaft und Energie 2016; Evans 2011; Rochet und Tirole 2006): In Plattformökonomien wird einer Personengruppe – z. B. Inhaber von Autos oder Schlafmöglichkeiten – der Zugang zu einer anderen Personengruppe – Mobilitäts- oder Schlafwillige – ermöglicht. Zentrales und damit konstituierendes Element dabei sind elektronisch vorliegende Datenbestände, deren elektronische Prozessierungsmöglichkeit sowie die nutzerfreundliche Handhabung der Angebote, in der Regel in Form von Apps auf mobilen Endgeräten. Über die so erreichbare Ortsunabhängigkeit und Ubiquität der eigentlichen Vermittlungsleistung entsteht die große Schlagkraft dieses Modells im Gegensatz zu etablierten Maklermodellen (Schwarze Bretter, Büros. Messen).⁹ Das Beispiel verdeutlicht nochmals die Relevanz des Zusammenspiels von Daten und Algorithmen (hier in Form einer nutzungsfreundlichen App).

Dorfer (2016, S. 315ff.) arbeitet literaturbasiert drei Merkmale heraus, die für datenzentrische Geschäftsmodelle konstituierend sein sollen: Sie bieten (1) eine kommerzielle Marktleistung an, die über Internettechnologien realisiert wird. Sie sind damit ein Sub-Typ von E-Business-Geschäftsmodellen, sind aber insofern spezifisch, als ihr (2) Leistungsangebot aus der Umwandlung von Daten in zweckdienliche Information besteht (*Datenintermediation*). Das Leistungsangebot adressiert (3) mehrseitige interdependente Märkte, deren Kundengruppen über mehrseitige Netzwerkexternalitäten verbunden sind. Literatur- und fallstudienbasiert ermittelt die Autorin (324ff.) dann in zwei Iterationen sechs Klassen von datenzentrischen Geschäftsmodellen. Darin sind als konkrete Ausprägungen z. B. Terminfindungstools, Bewertungsplattformen, Content-Aggregatoren, Anzeigenportale oder Social-Networking-Sites enthalten. Aus der Arbeit von Otto et al. (2016) zum Industrial Data Space lassen sich darüber hinaus Anregungen zu Rollen in und relevanten Gestaltungsbereichen von datenzentrierten Geschäftsmodellen entnehmen. Weiterhin liefern die Arbeiten von Bründl et al. (2016) sowie Palmetshofer et al. (2017) zum Handel mit persönlichen Daten Erkenntnis über die Funktionsweise von Datenmärkten. Auch Kitchin (2014, S. 42) beschäftigt sich unter der Bezeichnung *Data Broker* mit Akteuren, die mit persönlichen Daten handeln, oder Wert für Nachfrager erzeugen, in dem sie Daten aus heterogenen Beständen integrieren, ebenso Nie und Han (2018). Ältere Arbeiten mit deutlich technischer Fokussierung sammeln sich um die Mitte der 2000er Jahre (z. B. Vouros et al. 2004; Tjoa et al. 2006).

Mit Daten als Produktionsfaktor der Domäne *Publishing Industry* bzw. *Medien- und Kommunikationswirtschaft* beschäftigen sich die Arbeiten von Hagenhoff (2015-2020); Eble und Hoch (2015-2020); Pellegrini (2012, 2014, 2017) und das Sonderheft Big Data and Media Management des *International Journal on Media Management* (Heft 1 2016).

⁹ Zur Zentralisierung von Informationen und frühen Formen von Informationsintermediäten vgl. Tantner (2011) oder Krajewski (2012).

4 Digitalisierung

4.1 Begriff und Konzeptualisierung

Digital bedeutet »stufenweise« oder »zu den Fingern gehörend« (Pfeifer 1993b; Peters 2016). Abstrakter formuliert bedeutet digital abzählbar (eben mit den Fingern). Digital sind demnach alle Techniken, die mit einer abzählbaren – und das bedeutet festgelegten – Menge an Entitäten arbeiten. Alle gewünschten Aussagen oder Zustände werden durch Kombination einer begrenzten Menge an Entitäten, wie z. B. Zeichen, erzeugt. Das Analoge ist gekennzeichnet durch Dichte: Zwischen zwei beliebige Entitäten passt immer noch eine weitere dazwischen, so dass Werte innerhalb eines definierten Bereichs jede beliebig kleine Abstufung annehmen können. Es entstehen kontinuierliche Übergänge zwischen zwei Entitäten. In dieser Logik sind alle Schriftsysteme digitaler Art (z. B. Schäfer 2004; Cramer 2011, S. 9f.¹⁰). Sie arbeiten mit einer ggf. großen, aber dennoch abzählbaren Menge an einzelnen Elementen in Form von z. B. standardisierten Graphemen oder tastbaren Erhebungen (Brailleschrift), die zu Aussagensystemen kombiniert werden. Zwischen zwei Elemente passt nichts dazwischen, außer eben *Abstand* als Gegenteil zum Zustand *dicht*. Besonders gut kombiniert werden können informationsarme Zeichen, also solche Zeichen, die nur einen geringen eigenen Informationsgehalt in sich tragen, wie es bei unserer lateinischen Alphabet der Fall ist. Von solchen Zeichen benötigt man nur wenige (also einen geringen Vorrat), während informationsreiche Zeichen einen großen Vorrat unterschiedlicher Elemente erfordern, wie z. B. im Fall ideographischer oder logographischer Schriftsysteme. Die Idee, einen zu druckenden Text in seine überschaubaren Buchstabenbestandteile zu zerlegen und mittels einzelner Lettern zu setzen, ist eine Ausprägung von Digitalisierung wie sich eine bessere kaum finden lässt: Gutenberg war mit seiner »abzählbaren Kombinatorik« (Kittler 1993, S. 180) früh im Digitalgeschäft unterwegs; und war aufgrund der Informationsarmut der Zeichen des lateinischen Alphabets damit erfolgreicher als die frühen Asiaten, deren Schriftsysteme mit gehaltvollen Zeichen sich als typografisch schlecht handhabbar erwiesen haben. Ein Extrembeispiel für sehr informationsarme und damit kombinationsfreundliche Zeichen ist das polyadische Zahlensystem zur Basis 2, dessen Zeichenvorrat auf zwei Elemente (Null, Eins) begrenzt ist: »In unübertreffbarer Zeichenökonomie taucht das Bit einzig als Gegensatz seiner eigenen Abwesenheit auf« (Kittler 1995, S. 105). Aus der Kombination von z. B. acht informationsarmen Bits der Ausprägung Null und Eins kann ein Zeichen des lateinischen Alphabets entstehen (B = 01000010).

Die Idee, Informationen mit genau zwei Ziffern darzustellen ist keine Erfindung der Gegenwart. Bereits Leibniz hat das griechische Alphabet so codiert (Ernst 2004, S. 54). Darüber hinaus lassen sich aber auch Operationen (i.S.v. Algorithmen) mit einer ziffernmäßigen Codierung durchführen. Damit liegt eine Universalcodierung von konzeptionell großer Mächtigkeit vor: Information (oder Daten) und Operatoren zu ihrer Verarbeitung sind in dieser Codierung in identischer, ziffernmäßiger Form repräsentiert (Werber 2004,

¹⁰ Cramer (2011, S. 10) wirft den Kulturpessimisten, die das Internet für den Niedergang des Texts verantwortlich machen, eine »Mischung aus Vulgär-McLuhan und technischer Ignoranz« vor.

S. 86). Ziffernmäßig ist dabei nicht gleichzusetzen mit binär, auch z. B. eine ternäre Codierung mit minus eins, null und eins wäre denkbar. Das Binäre hat sich durchgesetzt, weil dieser Code einfach elektrifizierbar ist, und genau hierin liegt die faktische Mächtigkeit: binär codierte Informationen und Operationen lassen sich konzeptionell auch über einen Papierspeicher (Turing-Maschine, Liebig 2011) nutzen; elektrifiziert entsteht aber der Vorteil der Unabhängigkeit von spezifischen Techniken und Materialitäten, wie Beschreibstoffen oder unterschiedlichen mechanischen Teilen (»General Purpose Technology«, Helpman 1998).

Elektrifizierung schafft zwei weitere Potenziale. Sie ermöglicht (1) eine einfachere Automatisierung. Automatisch sind solche Techniken, die Operationen selbsttätig durchführen, solange, bis entweder ein definiertes Ziel erreicht (siehe hierzu *Terminiertheit*) oder keine weitere Energie mehr zugeführt wird. Um eine Kette von binärcodierten Operationen mit einer physischen¹¹ Maschine automatisch abzuarbeiten zu können, kann – muss nicht – ein Computer im heutigen Verständnis benutzt werden: das ist ein Gerät mit Prozessor, bestehend aus Rechenwerk und Steuerwerk, Speicher sowie Eingabe- und Ausgabeschnittstelle. Diese so genannte Von-Neumann-Architektur speicherprogrammierter Maschinen ist in ihrer Beschaffenheit völlig unabhängig vom konkreten, zu lösenden Problem. In der Tendenz sind diese Maschinen bis dahin immer leistungsfähiger, kleiner und billiger geworden (zu den Entwicklungsgrenzen Vowinkel 2017). In Software manifestierte Algorithmen liefern das automatisiert abzuarbeitende Programm.

Elektrifizierung erleichtert (2) die Vernetzung. Vernetzung im sozialen Sinne erfordert Beziehungen zwischen Akteuren. Nach (Castells 2009, S. 20) sind soziale Netzwerke kommunikative Strukturen. Im Falle räumlicher Distanz zwischen Akteuren sind Kommunikationsmedien sowie physische Infrastrukturen und Organisationsleistungen zur Pflege der Beziehungen erforderlich, bei (Behringer 2005, 2003, S. 63) am Beispiel des Postwesens für die Frühe Neuzeit beschrieben. Winkler (2004, S. 212) argumentiert, dass im Laufe der Medien- und Kommunikationsgeschichte Zeichenträger immer kleiner und leichter geworden sind (Prozess der Immaterialisierung), was dem Umstand zugutekommt, dass Zeichen zwecks Kommunikation zirkulieren wollen (auch Reischer 2006, S. 18). Der konzeptionelle Endpunkt der Leichterwerdung und Erleichterung der Zirkulation liegt in ihrer Elektronifizierung; Osterhammel (2011, Pos. 1578f. / Kap. 5) spricht in Bezug auf die Verbreitung der elektrischen Telegrafie im 19. Jahrhundert von der *Verkabelung der Welt*. Ein elektronisches Universalgerät wie oben beschrieben kann am Senden und Empfangen der leichten Zeichen über seine Ein- und Ausgabeschnittstellen teilhaben, damit wird es ein äußerst potenter »Knoten im Netz der Zeichenzirkulation« (Winkler 2004, S. 213).

4.2 Ausprägungen von Digitalisierung 1: Wertschöpfung

Aus einer Wertschöpfungsperspektive betrachtet kann Digitalisierung im Sinne von Elektronifizierung von bzw. Einsatz von Software in Prozessen und Arbeitsabläufen stattfinden. Dabei geht es um Ziele wie Effizienzsteigerung (z. B. Verarbeitung von mehr Geschäftsvorfällen in gleicher Zeit), Effektesteigerung (z. B. höhere Qualität da größere Präzision), Vermeiden von unangenehmen Arbeiten (z. B. Schmutz, Gefahr, oder so genannte

¹¹ Die oben erwähnte Turing-Maschine ist ein rein theoretisches Denkmodell, das ohne physisches Artefakt auskam und auskommt.

›Routine‹) oder Ersatz von bisherig genutzten, aber knapper werdenden Ressourcen (z. B. menschliche Arbeitskraft in erforderlicher Menge): »Wir Menschen verschwenden zu viel Intelligenz für Routineaufgaben. Irgendwann werden wieder weniger Menschen auf der Erde leben. Dann wird mehr Unterstützung durch Technik gebraucht. Außerdem verschwenden wir viel zu viele natürliche Ressourcen. Künstliche Intelligenz wird dabei helfen, die Umwelt zu schonen. Das beginnt schon mit dem Energieverbrauch zu Hause [...]« (Floridi 18.07.2016). Elektronifiziert werden kann auch das Ergebnis von Prozessen, also die Leistung oder das Gut, welches aus einem Transformationsprozess von Input zu Output resultiert. Güter, verstanden als Bündel von Eigenschaften, bestehen dann sehr grob aus Bits und nicht (mehr) aus Atomen (Negroponte 1995, 1996; Schmitt 2019).

Die nachstehende Abbildung führt die Dimensionen *Prozess* mit den groben Aktivitäten *Leistung erzeugen* und *Raumüberbrückung* sowie *Objekt* zusammen. Der Grad der Digitalisierung der Prozesse der Leistungserzeugung kann zwischen 0 % (keine Elektronifizierung) und 100 % (vollständig elektronifiziert) liegen. In der Überbrückung des Raums zwischen dem Ort der Leistungserstellung und dem Ort der Verwendung der Leistung (Konsum, Weiterverarbeiter) liegt eine dichotome Ausprägung vor: entweder Objekte werden mithilfe physisch-materieller Infrastruktur räumlich bewegt (Flugzeug, Schiff, LKW, Straße) oder mithilfe digitaler Infrastrukturen durch Versenden von Dateien, die das zu bewegende Objekt darstellen oder repräsentieren. Das Objekt selber kann in digitaler oder in physisch-materieller Form vorliegen. Szenario 1 stellt das traditionelle Handwerk dar, exemplarisch hier der Druck mit Bleiletern. In Szenario 2 werden Prozesse mittels IT unterstützt, aber nicht gänzlich bitbasiert abgewickelt, was vor allem daran liegen kann, dass das zu behandelnde Gut in physischer Form vorliegt (Steuerung von Druckstraßen, Lagerbewirtschaftung in der Medienlogistik) oder nach wie vor zwingend Leistungsbeiträge erforderlich sind, die unmittelbar nur von Menschen erbracht werden können (redaktionelle Leistung im crossmedialen Publishing). In Kapitel 3.3 entspricht das den Phasen »Daten als Befähiger von Prozessen bzw. Produkten«. Szenario 3 erlaubt eine (nahezu) vollständige Automatisierung von Prozessschritten. Als konkretes Beispiel kann hier der automatische Satz von Text- und Bildarrangements dienen, der nach wohldefinierten Layoutregeln die First Copy eines Druckwerks generiert. Ebenfalls fallen automatisch erzeugte Texte oder Bilder in dieses Szenario (hierzu z. B. (Napoli 2014b, 2014a)). Mit Szenario 4 ist die dezentrale Fertigung verbunden, in der eigentlich physische Gegenstände zunächst als Datensatz abgebildet und über Kommunikationsinfrastrukturen an einen anderen Ort transportiert werden, um dann dort hergestellt zu werden. Der 3D-Druck ist ein sehr aktuelles Beispiel hierfür, aber auch der dezentrale Digitaldruck von inhaltstragenden Medien kann hier genannt werden. Ein fast schon skurril wirkendes Beispiel ist die Produktion der enormen Auflage von Barack Obamas Memoiren (»A promised Land«): Die 3 Millionen Exemplare Erstauflage wurden zu 50 % in den USA gedruckt und zu 50 % in Deutschland, gleichwohl ist der größte Teil dieser Auflage für den US-amerikanischen Markt gedacht, so dass das dezentral produzierte, physische Erzeugnis mit dem Schiff zurücktransportiert wurde. Szenario 5 stellt einen nahezu voll elektronifizierten Prozess dar, in welchem rein elektronische Güter hervorgebracht werden, die elektronisch ausgeliefert werden.

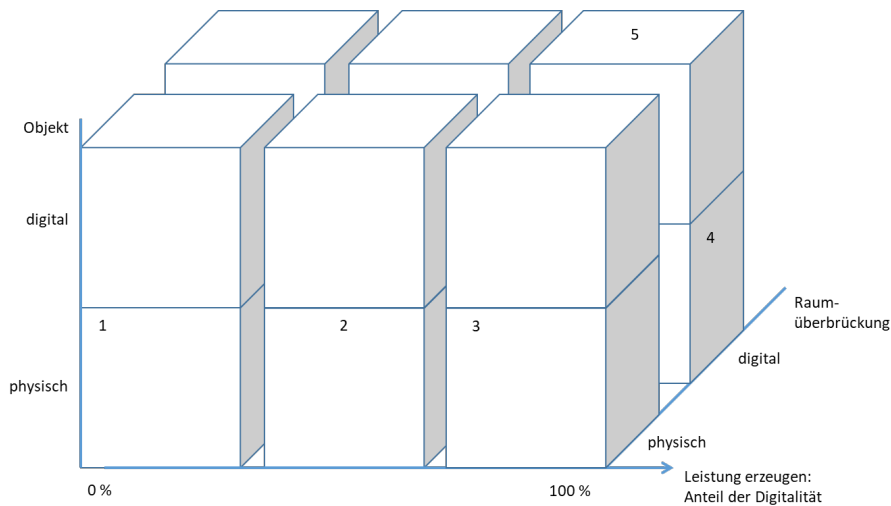


Abbildung 4: Ausprägungen von Digitalisierung in der Wertschöpfung

4.3 Ausprägungen von Digitalisierung 2: Relation zwischen Entitäten und Bits

Ausprägungen von Digitalisierung können auch anhand der Relation dekliniert werden die besteht zwischen Entitäten einerseits und den Bits andererseits, die als elektronischer Datenbestand die Entitäten repräsentieren oder selber sind und Verarbeitungen oder Bearbeitungen ermöglichen. In dieser Logik können drei Digitalisierungsausprägungen identifiziert werden:

Entitäten sind grundsätzlich abstrakte (z. B. Prozesse, Kunden, Bürger), künstliche (z. B. Produkte, Kunstwerke) oder natürliche (z. B. Bäume, Moleküle, konkrete Menschen) Gegenstände, Zustände, Subjekte oder Konzepte der Realwelt, die in ihren Eigenschaften beschrieben und erfasst werden können. Diese Beschreibungen liegen in Form von Zeichenketten vor, die elektronisch als Bits abgelegt und organisiert werden, z. B. in Datenbanken. Die Relation besteht darin, dass Entitäten durch Bits *repräsentiert* werden durch die Deklaration adäquater Attribute, die Entitäten aber auch ohne diese Repräsentation bestehen. Die Attributwerte können elektronisch *verarbeitet* werden, z. B. in Szenarien der Prozesssteuerung. Je strukturierter (eindeutige Attributierung) diese Beschreibungen vorliegen, desto einfacher und auch eindeutiger sind sie automatisiert zu verarbeiten. Die nachstehende Abbildung zeigt die Beschreibung einer Bibliothek (Notation Zustandsdiagramm in UML, dieses und weitere Beispiele bei Hagenhoff 2015-2020, 2016), in welcher Titalexemplare von Büchern als konkrete Entitäten bereitgestellt werden und ausgeliehen werden können.

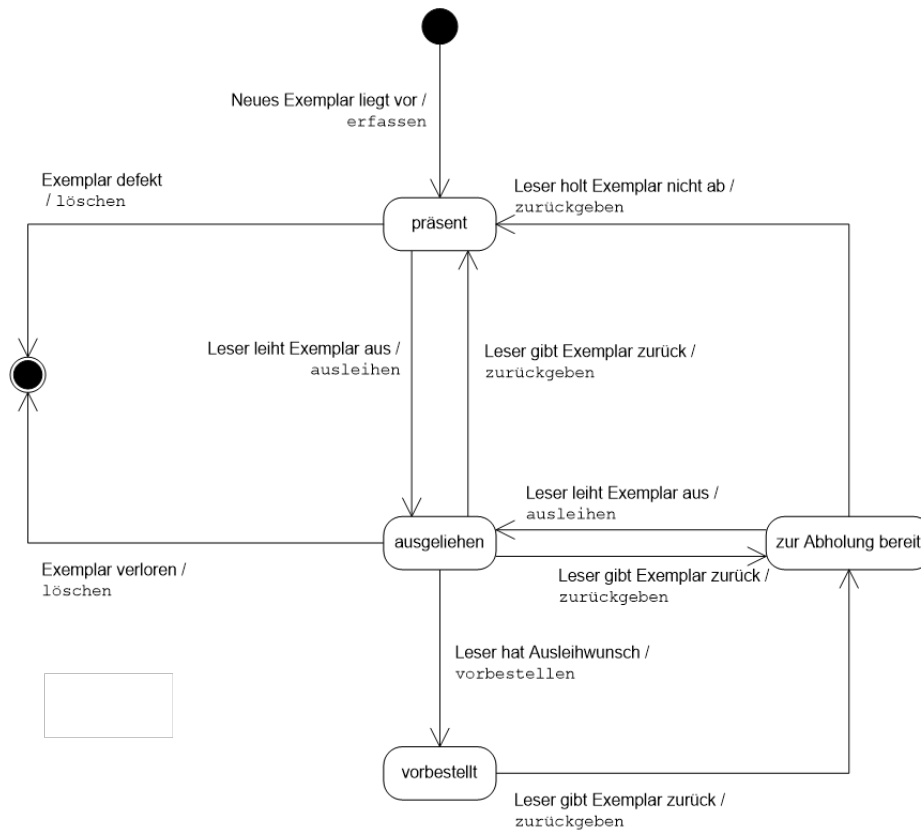


Abbildung 5: Beschreibung einer Bibliothek

Einige der genannten Entitätsarten können digital *abgebildet* werden, indem sie fotografiert oder 2- oder 3-dimensional gescannt werden. Die digitalen Abbildungen liegen als Dateien verschiedener Formate vor, sie sind technische Kapseln. Die Relation besteht darin, dass die digitalen Abbilder als *Surrogate* für die eigentlichen Entitäten stehen und genutzt werden. Einher mit diesem Typ geht als komplexes Aufgabenfeld das Asset-Management, das sich der (Langzeit-)Archivierung sowie dem Auffinden, Anzeigen und Bündeln der Dateien widmet, dargestellt in Abbildung 6 – links die Oberfläche der prometheus-Datenbank, rechts Architektur eines Asset-Management-Systems (Austerberry 2006). Die Surrogate wiederum werden mit Hilfe von deskriptiven, technischen und administrativen Metadaten (Rehbein 2017) beschrieben.

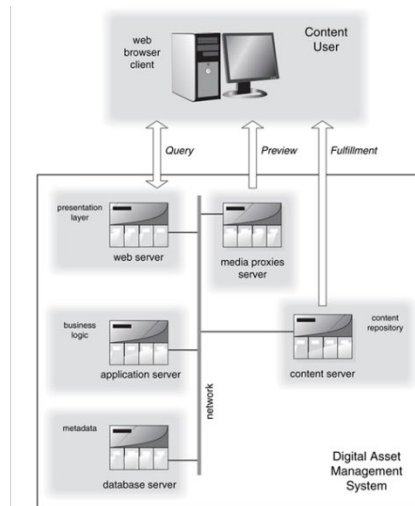


Abbildung 6: Asset Management

Entitäten können *nativ digital* vorliegen. Bits sind von vornherein das ›Material‹ aus dem ›Etwas‹ erzeugt wird, z. B. Texte, Bilder, Töne oder Modelle (siehe auch Kapitel 3.3). Nicht zu jeder nativ digitalen Entität gehört auch zwangsläufig eine reale Entität, so existieren z. B. E-Books ohne physisches Pendant, nicht mehr existente Objekte können rekonstruiert werden (Abbildung 7, Kohle 2017) oder Denkbare, Mögliches oder Wahrscheinliches kann simuliert werden (hierzu z. B. Saam 2017). Nativ digitale Entitäten sind in ihrer Existenz somit nicht zwingend anhängig von realweltlich existierenden Entitäten. Sie können *umfangreich manipuliert* werden, was häufig auch ihr Sinn ist (z. B. Parameterveränderung in Modellen). Ein Beispiel für umstrittenes Potenzial von Manipulation ist das Projekt SiDiM¹² – Sichere Dokumente durch individuelle Markierung – des Fraunhofer-Instituts, das von der Buchbranche mit Empörung aufgenommen wurde (Sine nomine 2013).

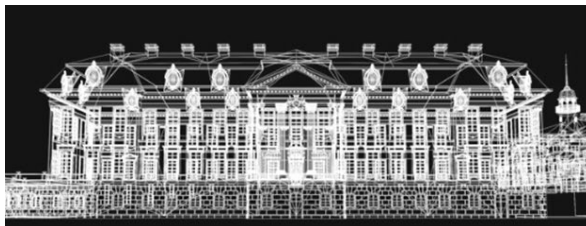


Abbildung 7: Rekonstruktionen (Drahtgittermodell links und texturiertes Modell rechts)

¹² <https://www.sit.fraunhofer.de/de/angebote/projekte/sidim/> [28.11.2020]

5 Versuch einer Synopse und Identifikation von Synapsen

5.1 Was ist jetzt eigentlich neu?

Binärcodierung, Elektrifizierung, die Von-Neumann-Universalarchitektur speicherprogrammierter Maschinen sowie deren Miniaturisierung und Verbilligung ergeben in Kombination eine ubiquitäre, unspezifische sowie potenziell vernetzte Technik, die Daten mit Hilfe von in Software manifestierten Algorithmen automatisch verarbeitet und die in zahlreiche Lebensbereiche auch jenseits von Büro und Fabrik eingedrungen ist. Ein differenzierter Diskurs bei Kübler (2005, S. 26ff.) zeigt, dass das so Beschriebene vor nicht allzu langer Zeit noch als *Informationsgesellschaft* oder *Informatisierung* bezeichnet wurde (vgl. auch Löffelholz und Altmeppen 1994).

Somit stellt sich die Frage nach der Differenzqualität des heutigen *Zeitalters der Digitalisierung* zum *Informationszeitalter*. In den letzten ca. 10 Jahren hat es zwei Entwicklungsschübe gegeben. Zum einen sind in Bezug auf Hardware extreme Fortschritte erzielt worden, vor allem die mobilen Endgeräte haben seit 2007 erhebliche Veränderungen erfahren. Sie werden als unabhkömmliches LiveStyle-Produkt überall mit hingenommen. Aber auch andere Arten von Hardware, wie z. B. Sensoren, sind mittlerweile so miniaturisiert und trotzdem leistungsstark, dass sie nahezu überall eingebaut werden können (»The Invisible Computer«, Norman 1999). Hardware ist somit also wirklich ubiquitär. Zum anderen hat eine *Miniaturisierung und Verbesserung der Benutzbarkeit von Software* stattgefunden. Damit ist gemeint, dass Software nicht mehr nur in Form sperriger und teurer Funktionsmonster mit schlechter Usability (vor allem in Bürouräumen) in Erscheinung tritt, sondern sich in Form relativ funktionsisolierter¹³, leicht bedienbarer Apps präsentiert, mit der viele Alltagsherausforderungen¹⁴ schnell gemeistert werden können. Der Paradigmenwechsel bei Hard- und Software kommt überdeutlich zum Ausdruck in dem Phänomen *Bring your own Device*, auch *Consumeration* genannt (Kohne et al. 2015): Mobile Endgeräte wie Smartphones und Tablets, und zwar am besten die eigenen, mit nützlicher, einfach zu bedienender Software möchten Mitarbeiter nun auch im Unternehmen nutzen.

Aus diesen zwei Gründen ergibt sich zum einen eine potenzielle Totalvernetzung von menschlichen bzw. technischen Akteuren (»Verkabelung der Welt« revisited). Zum anderen entstehen automatisiert Daten höchst unterschiedlicher Beschaffenheit, auch aus bisher nicht oder nur sehr aufwändig erschließbaren Quellen, vor allem Interaktionsdaten (hierzu z. B. Bensberg und Weiß 1999 oder Kaspar 2006) machen die Spezifität von Big Data im Vergleich zu herkömmlicher Business Analytics aus, und sie sind es, die die gesellschaftliche und gesellschaftswissenschaftliche Relevanz begründen. Die extrem gestiegene Rechenleistung ermöglicht eine neuartige Verarbeitung dieser Datenmengen in Form algorithmischer Modellierung und Induktion. Hiermit wird das Potenzial geschaffen, dass sehr komplexe, variablenreiche Probleme automatisiert genauer, verlässlicher oder

¹³ Zur Funktionsgröße von Software vgl. Fahsel et al. (2017, S. 39).

¹⁴ Die Gesellschaft für Informatik hat die Allgegenwärtige Mensch-Maschine-Interaktion zu einer Grand Challenge der Informatik erklärt (<https://gi.de/grand-challenges/>), hierzu auch bereits Norman (1999).

schneller bearbeitet werden könnten als ggf. dieses der Mensch mit seinem (Teil-)beschränkten Erkenntnisapparat leisten kann (Pietsch und Wernecke 2017, S. 17ff.).

5.2 Und nun?

Ausgangspunkt dieses Aufsatzes war die Feststellung eines Desiderats in Bezug auf die Fassungen der Konzepte Algorithmen, Daten sowie Digitalisierung. Die Ausführungen haben gezeigt, dass trotz Versuchen an Definitionen und Konzepterläuterungen etliche Unschärfen festzustellen sind, zudem für gründliche Einschätzungen der Potenziale und Herausforderungen tiefergehendes Wissen erforderlich ist:

Algorithmus als Begriff sowie Konzept ist recht klar zu fassen. Dennoch wäre es adäquater mit dem Begriff *Software* zu arbeiten, denn: Software ist die *automatisierte* Variante des Problemlösers Algorithmus, und das stellt eine spezifische und die gesellschaftlich sowie gesellschaftswissenschaftlich relevante Eigenschaft dar. Gegebenfalls werden durch den Versuch etwas zu automatisieren geringe Rationalitäten, Unlogiken, Inkonsistenzen oder fehlende Begründungen bisheriger menschlicher Routinen und Entscheidungsprozesse aufgedeckt, und vielleicht löst genau das die Aufregung in Wissenschaft und Feuilleton aus. Um die Potenziale und Herausforderungen von Automatisierungen für die medienvermittelte Kommunikation sowie Organisationen und Institutionen beurteilen zu können, ist jedoch tiefergehendes Wissen über die Arten von Algorithmen sowie vor allem ihre Grenzen als auch die Beschaffenheit von Software erforderlich. Eine Zusammenarbeit mit Informatikern ist für solche Fragestellungen dringend erforderlich, gleichzeitig lernen Informatiker die Anwendungsfälle und Domänen besser kennen, in denen ihre Technologien (sinnvoll) eingesetzt werden sollen.

Der Begriff und das Konzept Daten ist, wie die Ausführungen deutlich gezeigt haben, sehr unscharf, und der jeweils verwendete Begriff für das gefühlt gleiche offenbar stark von Moden abhängig (Daten, Datengetriebene Zeiten, Datenintermediäre; Informationen, Informationsgesellschaft, Informationsintermediäre; Wissen, Wissensgesellschaft). Mit Bezug auf die Medien- und Kommunikationsforschung soll hier – vorläufig – abschließend zwischen zwei grundverschiedenen Verwendungszwecken von Daten unterschieden werden (ähnlich auch Napoli 2014b): Kommunikate bestehen aus Daten (Hagenhoff 2015-2020; Owens 2011) in verschiedenen Erscheinungsformen (Bild, Text, Ton), die unterschiedlich gut automatisiert be- und verarbeitet werden können. Daten heißen in diesem Zusammenhang ggf. eher *Zeichen*, Voß (2013a, 2013b) hingegen setzt sie wohlbegründet mit *Dokument* gleich. Um Forschung hierzu zu bewerkstelligen ist eine Zusammenarbeit mit Semiotikern, Sprachwissenschaftlern, Computerlinguisten, Informationswissenschaftlern, Bildwissenschaftlern und weiteren Disziplinen sinnvoll, die mit Texten (in sehr weitem Begriffsverständnis), Bild und Audio umgehen können. Daten dienen wie in jeder Branche so auch in der Medien- und Kommunikationswirtschaft auch der Beschreibung von Zuständen und Aktivitäten, und diese Beschreibungen werden benötigt, um Entscheidungen treffen zu können. Mit welcher Art von Daten überhaupt was wie sinnvoll analysiert werden kann und ob das hinsichtlich stets zu verhandelnder Ziele und Werte dann adäquat ist (hierzu Orwat et al. 2010) kann erarbeitet werden zusammen mit Techniksoziologen, Wirtschaftsinformatikern, Juristen und Ethikern.

Literaturverzeichnis

- Altenhöner, Florian (2008): Kommunikation und Kontrolle. Gerüchte und städtische Öffentlichkeiten in Berlin und London 1914/1918. München, Berlin.
- Benenson, Zinaida; Berger, Frederik; Cherepantsev, Anatolij; Datsevich, Sergey; Del Medico, Katharina; Do, Long et al. (2019): Analyse verbreiteter Anwendungen zum Lesen von elektronischen Büchern. Technical Reports Nr. CS-2019-06. Hg. v. Department Informatik. Friedrich-Alexander-Universität Erlangen-Nürnberg. Erlangen.
- Beniger, James Ralph (1986): The control revolution. Technological and economic origins of the information society. 5. (E-Book-Version) Auflage. Cambridge (Mass.).
- Christmann, Ursula (2015): Lesen als Sinnkonstruktion. In: Ursula Rautenberg und Ute Schneider (Hg.): Lesen - Ein interdisziplinäres Handbuch. Berlin, S. 169–184.
- Cordón García, José Antonio (2013): Social reading. Platforms, applications, clouds and tags. Oxford, UK.
- Drotner, Kirsten (1999): Dangerous Media? Panic Discourses and Dilemmas of Modernity. In: Paedagogica Historica: International Journal of the History of Education 35 (3), S. 593–619.
- Dürscheid, Christa (2007): Private, nicht-öffentliche und öffentliche Kommunikation im Internet. In: Neue Beiträge zur Germanistik 6 (4), S. 22–41.
- Austerberry, David (2006): Digital asset management. 2. Auflage. Oxford.
- Ballsun-Stanton, Brian (2010): Asking about data. Experimental philosophy of Information Technology. In: Franz Ko (Hg.): Proceedings of the 5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT) Seoul. Piscataway NJ, S. 119–124.
- Behringer, Wolfgang (2003): Im Zeichen des Merkur. Reichspost und Kommunikationsrevolution in der Frühen Neuzeit. Göttingen.
- Behringer, Wolfgang (2005): »Von der Gutenberg-Galaxis zur Taxis-Galaxis«. Die Kommunikationsrevolution: Ein Konzept zum besseren Verständnis der Frühen Neuzeit. In: *Historische Zeitschrift. Beihefte* 41 (o. Nr.), S. 39–54.
- Beniger, James Ralph (1986): The control revolution. Technological and economic origins of the information society. 5. (E-Book-Version) Auflage. Cambridge (Mass.).
- Bensberg, F.; Weiß, T. (1999): Web Log Mining als Marktforschungsinstrument für das World Wide Web. In: *Wirtschaftsinformatik* 41 (5), S. 426–432.
- Bentele, Günter; Brosius, Hans-Bernd; Jarren, Otfried (Hg.) (2013): Lexikon Kommunikations- und Medienwissenschaft. 2., überarb. und erweiterte Auflage. Konstanz.
- Boockmeyer, Arne; Fischbeck, Philipp; Neubert, Stefan (2017): Fit fürs Studium - Informatik. Bonn.
- Brosche, Kolja (2014): Echtzeit-Daten werden Treibstoff digitaler Werbung. In: Oliver Busch (Hg.): Realtime Advertising: Digitales Marketing in Echtzeit: Strategien, Konzepte und Perspektiven. Wiesbaden, S. 187–197.
- Bründl, Simon; Matt, Christian; Hess, Thomas (2016): Daten als Geschäft — Rollen und Wertschöpfungsstrukturen im deutschen Markt für persönliche Daten. In: *Wirtschaftsinformatik & Management* o. Jg. (6), S. 66–71.

- Bundesministerium für Wirtschaft und Energie (2016): Grünbuch Digitale Plattformen. Berlin.
- Bundesministerium für Wirtschaft und Energie (2020): GWB-Digitalisierungsgesetz. Online verfügbar unter <https://www.bmwi.de/Redaktion/DE/Artikel/Service/Gesetzesvorhaben/gwb-digitalisierungsgesetz.html> [28.11.2020]
- Camenisch, Jan; Fischer-Hübner, Simone; Hansen, Marit; Tsormpatzoudi, Pagona; Dimitrova, Diana; Schroers, Jessica; Kindt, Els (Hg.) (2015): Privacy by Design – The Case of Automated Border Control. Privacy and Identity Management for the Future Internet in the Age of Globalisation. Berlin.
- Carr, Nicholas G. (2008): The big switch. Rewiring the world, from Edison to Google. New York City.
- Castells, Manuel (2009): Communication power. Oxford, UK.
- Cramer, Florian (2011): Exe.cut(up)able statements. Poetische Kalküle und Phantasmen des selbstaufführenden Texts. Paderborn.
- Dijkstra, Edsger Wybe (1972): The humble programmer. In: *Communications of the ACM* 15 (10), S. 859–866.
- Dorfer, Laura (2016): Datenzentrische Geschäftsmodelle als neuer Geschäftsmodelltypus in der Electronic-Business-Forschung. Konzeptionelle Bezugspunkte, Klassifikation und Geschäftsmodellarchitektur. In: *ZfbF* 68 (3), S. 307–369.
- Drösser, Christoph (2016): Total berechenbar? Wenn Algorithmen für uns entscheiden. München.
- Eble, Michael; Hoch, Julian (2015-2020): Datenbanken, Data Warehousing & Data Analytics. In: Jan Krone und Tassilo Pellegrini (Hg.): Handbuch Medienökonomie. Berlin, Aufsatz online first 2016.
- Ernst, Wolfgang (2004): Den A/D-Umbruch aktiv denken – medienarchäologisch, kulturtechnisch. In: Alexander Böhnke und Jens Schröter (Hg.): Analog/Digital - Opposition oder Kontinuum? Zur Theorie und Geschichte einer Unterscheidung. Bielefeld, S. 49–66.
- Evans, David (Hg.) (2011): Platform Economics: Essays on multi-sides business. Competition Policy International.
- Fahsel, Jörn; Hagenhoff, Svenja; Heinold, Ehrhardt (2017): Publishing 4.0. Chancen, Anforderungen, Konzepte. Denkzeug 2017: Cross-, Hybrid-Media und Digital Content-Services. Hg. v. Institut für Buchwissenschaft, Universität Erlangen-Nürnberg und Heinold, Spiller und Partner Unternehmensberatung. Erlangen und Hamburg.
- Floridi, Luciano (2005): Is Semantic Information Meaningful Data? In: *Philos Phenomenol Res* 70 (2), S. 351–370.
- Floridi, Luciano (2009): Philosophical Conceptions of Information. In: Giovanni Sommaruga (Hg.): Formal Theories of Information: From Shannon to Semantic Information Theory and General Concepts of Information. Berlin, Heidelberg, S. 13–53.
- Floridi, Luciano (2010): Information. A very short introduction. Oxford.
- Floridi, Luciano interviewt von Borchardt, Alexandra (18.07.2016): »Welche Welt wollen wir?«. Süddeutsche Zeitung Magazin »Künstliche Intelligenz«.
- Garey, Michael R.; Johnson, David S. (1979): Computers and intractability. A guide to the theory of NP-completeness. New York, NY.

- Gerpott, Torsten J. (2020): Mit Wattebäuschchen gegen Plattformriesen. In: *FAZ* (20.11.2020). Online verfügbar unter <https://zeitung.faz.net/faz/medien/2020-11-20/15a65b8521a590971448bd689b83a739/?GEPC=s9> [21.11.2020].
- Gillespie, Tarleton (2014): The Relevance of Algorithms. In: Tarleton Gillespie, Pablo J. Boczkowski und Kirsten A. Foot (Hg.): *Media technologies. Essays on communication, materiality, and society*. Cambridge (Mass.), S. 167–193.
- Glaß, Marco; Schmidt, Thorsten (2011): Aufbau und Weiterentwicklung der Datenqualitätsmessungen bei der Telekom Deutschland GmbH. In: *HMD* 48 (3), S. 56–63.
- Gumm, Hans-Peter; Sommer, Manfred (2008): *Einführung in die Informatik*. 8. Auflage. München.
- Hagel, John III; Rayport, Jeffrey F. (1999): The coming battle for customer information. In: *Harvard business review* 75.
- Hagenhoff, Svenja (2015-2020): Management von Content in den Kernprozessen der Produktion und Distribution. In: Jan Krone und Tassilo Pellegrini (Hg.): *Handbuch Medienökonomie*. Berlin, Paper online first 2015.
- Hagenhoff, Svenja (2016): Systems Analysis as a Basis for Designing IT in the Media Industry. In: Artur Lugmayr, Emilija Stojmenova, Katarina Stanoevska und Robert Wellington (Hg.): *Information Systems and Management in eMedia and Creative Industries*. Berlin, S. 207–222.
- Heesen, Jessica (Hg.) (2016): *Handbuch Medien- und Informationsethik*. Stuttgart.
- Helpman, Elhanan (Hg.) (1998): *General Purpose Technologies and Economic Growth*. Cambridge (Mass.)
- Hofstetter, Yvonne (2018): *Das Ende der Demokratie. Wie die künstliche Intelligenz die Politik übernimmt und uns entmündigt*. München.
- Jannidis, Fotis; Kohle, Hubertus; Rehbein, Malte (Hg.) (2017): *Digital Humanities: Eine Einführung*. Stuttgart.
- Kaspar, Christian (2006): *Individualisierung und mobile Dienste am Beispiel der Medienbranche. Ansätze zum Schaffen von Kundenmehrwert*. Göttingen.
- Keller, Rudi (1995): *Zeichentheorie. Zu einer Theorie semiotischen Wissens*. Tübingen.
- Kitchin, Rob (2014): *The Data Revolution. Big Data, Open Data, Data Infrastructures and Their Consequences*. London.
- Kittler, Friedrich A. (1993): Geschichte der Kommunikationsmedien. In: Jörg Huber und Alois Martin Müller (Hg.): *Raum und Verfahren*. Basel, Frankfurt am Main, S. 169–188.
- Kittler, Friedrich A. (1995): Von der Letter zum Bit. In: Horst Wenzel (Hg.): *Gutenberg und die Neue Welt*. München, S. 105–117.
- Kohle, Hubertus (2017): Digitale Rekonstruktion und Simulation. In: Fotis Jannidis, Hubertus Kohle und Malte Rehbein (Hg.): *Digital Humanities: Eine Einführung*. Stuttgart, S. 315–326.
- Kohne, Andreas; Ringleb, Sonja; Yücel, Cengizhan (Hg.) (2015): *Bring your own Device: Einsatz von privaten Endgeräten im beruflichen Umfeld – Chancen, Risiken und Möglichkeiten*. Wiesbaden.
- Kowalski, Robert (1979): Algorithm = Logic + Control. In: *Communications of the ACM* 22 (7), S. 424–436.

- Krajewski, Markus (2012): Ask Jeeves. Der Diener als Informationszentrale. In: Thomas Brandstetter, Thomas Hübel und Anton Tantner (Hg.): Vor Google. Eine Mediengeschichte der Suchmaschine im analogen Zeitalter. Bielefeld, S. 151–172.
- Krämer, Sybille (2005): »Operationsraum Schrift«. Über einen Perspektivenwechsel in der Betrachtung der Schrift. In: Gernot Grube, Werner Kogge und Sybille Krämer (Hg.): Schrift: Kulturtechnik zwischen Auge, Hand und Maschine. München, S. 23–57.
- Kübler, Hans-Dieter (2005): Mythos Wissensgesellschaft. Gesellschaftlicher Wandel zwischen Information, Medien und Wissen. Eine Einführung. Wiesbaden.
- Laudon, Kenneth C.; Laudon, Jane Price; Schoder, Detlef (2016): Wirtschaftsinformatik. Eine Einführung. 3., vollständig überarbeitete Auflage. Hallbergmoos.
- Liebig, Hans (2011): Konrad Zuse, Erfinder des Computers – im Vergleich mit Alan Turing und John v. Neumann. In: *Informatik-Spektrum* 34 (6), S. 553–564.
- Löffelholz, Martin; Altmeppen, Klaus-Dieter (1994): Kommunikation in der Informationsgesellschaft. In: Klaus Merten, Siegfried J. Schmidt und Siegfried Weischenberg (Hg.): Die Wirklichkeit der Medien: eine Einführung in die Kommunikationswissenschaft. Opladen, S. 570–591.
- Loleit, Simone (2004): »The mere digital process of turning over leaves«. Zur Wort- und Begriffsgeschichte Von »digital«. In: Alexander Böhnke und Jens Schröter (Hg.): Analog/Digital - Opposition oder Kontinuum? Zur Theorie und Geschichte einer Unterscheidung. Bielefeld, S. 193–214.
- Meyer-Wegener, Klaus (2017): Erstmal einfach alles speichern: Big Data als Aufgabe für die Informatik. In: Rudolf Freiburg (Hg.): D@tenflut. Erlangen, S. 99–115.
- Möller, Frederik; Spiekermann, Markus; Burmann, Anja; Pettenpohl, Heinrich; Wenzel, Sven (2017): Bedeutung von Daten im Zeitalter der Digitalisierung. Whitepaper. Hg. v. Michael ten Hompel, Michael Henke und Uwe Clausen. Fraunhofer-Institut für Materialfluss. Dortmund.
- Mowshowitz, Abbe (1984): Computers and the Myth of Neutrality. In: Frank L. Friedman (Hg.): Proceedings of the ACM 12th annual computer science conference on SIGCSE symposium. New York, NY, S. 85–92.
- Müller, Roland M.; Lenz, Hans-Joachim (2013): Business Intelligence. Berlin, Heidelberg.
- Napoli, Philip (2014a): Automated Media. An Institutional Theory Perspective on Algorithmic Media Production and Consumption. In: *Communication Theory* 24 (3), S. 340–360. DOI: 10.1111/comt.12039.
- Napoli, Philip (2014b): On Automation in Media Industries. Integrating Algorithmic Media Production into Media Industries Scholarship. In: *Media Industries Journal* 1 (1), S. 33–38.
- Negroponte, Nicholas (1995): Bits and Atoms. Publiziert auf *Wired* am 1. Januar. Online verfügbar unter <http://www.wired.com/1995/01/negroponte-30/> [28.11.2020].
- Negroponte, Nicholas (1996): Being digital. New York, NY.
- Nie, Yingzhi; Han, Xueping (2019): Research on consumers' protection in advantageous operation of big data brokers. In: *Cluster Computing* 22 (S4), S. 8387–8400.
- Norman, Donald A. (1999): The invisible computer. Why good products can fail, the personal computer is so complex, and information appliances are the solution. Cambridge (Mass.).

- Orwat, Carsten; Raabe, Oliver; Buchmann, Erik; Anandasivam, Arun; Freytag, Johan-Christoph; Helberger, Natali et al. (2010): Software als Institution und ihre Gestaltbarkeit. In: *Informatik-Spektrum* 33 (6), S. 626–633.
- Osterhammel, Jürgen (2011): Die Verwandlung der Welt. Eine Geschichte des 19. Jahrhunderts. München.
- Ottmann, Thomas; Widmayer, Peter (2017): Algorithmen und Datenstrukturen. 6. Auflage. Berlin, Heidelberg.
- Otto, Boris; Jürjens, Jan; Schon, Jochen; Auer, Sören; Menz, Nadja; Wenzel, Sven; Cirullies, Jan (2016): Industrial Data Space. Whitepaper. Hg. v. Fraunhofer-Gesellschaft. München.
- Owens, Trevor (2011): Defining Data for Humanists. Text, Artifact, Information or Evidence? In: *The Journal of Digital Humanities* 1 (1).
- Palmetshofer, Walter; Semsrott, Arne; Alberts, Anne (2017): Der Wert persönlicher Daten. Ist Datenhandel der bessere Datenschutz? Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen.
- Pellegrini, Tassilo (2012): Semantic Web in netzbasierten Unterhaltungsmedien. Bausteine für eine Metadaten-Ökonomie. In: Jörg Müller-Lietzkow (Hg.): Ökonomie, Qualität und Management von Unterhaltungsmedien. Baden-Baden, S. 253–276.
- Pellegrini, Tassilo (2014): Datenlizenzierung als Diversifikationstreiber in der Medienindustrie. In: Harald Rau (Hg.): Digitale Dämmerung. Die Entmaterialisierung der Medienwirtschaft. Baden-Baden, S. 267–280.
- Pellegrini, Tassilo (2017): Semantic metadata in the publishing industry – technological achievements and economic implications. In: *Electronic Markets* 27 (1), S. 9–20.
- Peters, Benjamin (2016): Digital. In: Benjamin Peters (Hg.): Digital keywords. A vocabulary of information society and culture. Princeton.
- Pfeifer, Wolfgang (1993a): Datum. In: Wolfgang Pfeifer (Hg.): Etymologisches Wörterbuch des Deutschen. Digitalisierte und von Wolfgang Pfeifer überarbeitete Version im Digitalen Wörterbuch der deutschen Sprache.
- Pfeifer, Wolfgang (1993b): Digital. In: Wolfgang Pfeifer (Hg.): Etymologisches Wörterbuch des Deutschen. Digitalisierte und von Wolfgang Pfeifer überarbeitete Version im Digitalen Wörterbuch der deutschen Sprache.
- Pietsch, Wolfgang; Wernecke, Jörg (2017): Einführung: Zehn Thesen zu Big Data und Berechenbarkeit. In: Wolfgang Pietsch, Jörg Wernecke und Maximilian Otte (Hg.): Berechenbarkeit der Welt? Philosophie und Wissenschaft im Zeitalter von Big Data. Wiesbaden, S. 13–35.
- Pólya, George (2010): Schule des Denkens. Vom Lösen mathematischer Probleme. Sonderausgabe der 4. Auflage. Tübingen.
- Rautenberg, Ursula (Hg.) (2015): Reclams Sachlexikon des Buches. 3., verbesserte Auflage. Stuttgart.
- Rehbein, Malte (2017): Digitalisierung. In: Fotis Jannidis, Hubertus Kohle und Malte Rehbein (Hg.): Digital Humanities: Eine Einführung. Stuttgart, S. 179–198.
- Reichert, Ramón (2018): Big Data als Boundary Objects. Zur medialen Epistemologie von Daten. In: Thorben Mämecke, Jan-Hendrik Passoth und Josef Wehner (Hg.): Bedeutende Daten: Modelle, Verfahren und Praxis der Vermessung und Verdatung im Netz. Wiesbaden, S. 17–33.

- Reischer, Jürgen (2006): Zeichen, Information, Kommunikation. Analyse und Synthese des Zeichen- und Informationsbegriffs. Regensburg.
- Rieder, Bernhard; Röhle, Theo (2012): Digital methods: Five challenges. In: David Berry (Hg.): Understanding digital humanities. London, S. 67–84.
- Rochet, Jean-Charles; Tirole, Jean (2006): Two-sided markets. A progress report. In: *The Rand journal of economics* 37 (3), S. 645–667.
- Rohweder, Jan P.; Kasten, Gerhard; Malzahn, Dirk; Piro, Andrea; Schmid, Joachim (2008): Informationsqualität — Definitionen, Dimensionen und Begriffe. In: Knut Hildebrand, Marcus Gebauer, Holger Hinrichs und Michael Mielke (Hg.): Daten- und Informationsqualität: Auf dem Weg zur Information Excellence. Wiesbaden, S. 25–45.
- Rosenberg, Daniel (2013): Data before the Fact. In: Lisa Gitelman (Hg.): »Raw data« is an oxymoron. Cambridge (Mass.), S. 15–40.
- Saam, Nicole J. (2017): What is a Computer Simulation? A Review of a Passionate Debate. In: *J Gen Philos Sci* 48 (2), S. 293–309.
- Schäfer, Jörgen (2004): Sprachzeichenprozesse. Überlegungen zur Codierung von Literatur in ‚alten‘ und ‚neuen‘ Medien. In: Alexander Böhnke und Jens Schröter (Hg.): Analog/Digital - Opposition oder Kontinuum? Zur Theorie und Geschichte einer Unterscheidung. Bielefeld, S. 143–168.
- Schmidt, Jan-Hinrik (2013): Social Media. 2. Auflage. Berlin.
- Schmidt, Siegfried J. (1994): Konstruktivismus in der Medienforschung. Konzepte, Kritiken, Konsequenzen. In: Klaus Merten, Siegfried J. Schmidt und Siegfried Weischenberg (Hg.): Die Wirklichkeit der Medien: eine Einführung in die Kommunikationswissenschaft. Opladen, S. 592–623.
- Schmitt, Bernd (2019): From Atoms to Bits and Back: A Research Curation on Digital Technology and Agenda for Future Research. In: *Journal of Consumer Research* 46 (4), S. 825–832.
- Scholz, Oliver R. (2004): Bild, Darstellung, Zeichen. Philosophische Theorien bildlicher Darstellung. 2., vollst. überarb. Auflage. Frankfurt am Main.
- Schöning, Uwe (2008): Ideen der Informatik. Grundlegende Modelle und Konzepte der Theoretischen Informatik. 3., korrigierte Auflage. Berlin/Boston.
- Schulz, Wolfgang; Dankert, Kevin (2016): Die Macht der Informationsintermediäre. Erscheinungsformen, Strukturen und Regulierungsoptionen. Hg. v. Friedrich-Ebert-Stiftung. Bonn.
- Schumann-Giesler, Ulrike (1997): Datenkonsistenz bei heterogener Datenspeicherung. Konzept und prototypische Realisierung. Wiesbaden.
- Sine nomine (2013): Eine sehr dumme Idee. In: *Buchreport Online* (21. Juni). Online verfügbar unter <https://www.buchreport.de/news/eine-sehr-dumme-idee/> [28.11.2020].
- Sjurts, Insa (Hg.) (2011): Gabler Lexikon Medienwirtschaft. 2., aktualisierte und erw. Auflage. Wiesbaden.
- Tantner, Anton (2011): Adressbüros im Europa der Frühen Neuzeit. Wien.
- Tempich, Christoph; Rieger, Volker (2007): Data-centric business models. Leveraging the hidden treasures of the telecom industry. Bonn.
- Tjoa, A. Min; Xu, Li; Chaudhry, Sohail S.; Calefato, Fabio; Colagrossi, Attilio; Gendarmi, Domenico et al. (Hg.) (2006): An Information Broker for Integrating Heterogeneous

- Hydrologic Data Sources: A Web Services Approach. Research and Practical Issues of Enterprise Information Systems. Boston (Mass.).
- Tschudi, Fabrice; D'Onofrio, Sara (2018): Wie wir durch unsere Denk- und Handlungsmuster beeinflusst werden. In: *HMD Praxis der Wirtschaftsinformatik* 55 (241), S. 467–471.
- Voß, Jakob (2013a): Describing data patterns. A general deconstruction of metadata standards. North Charleston.
- Voß, Jakob (2013b): Was sind eigentlich Daten? In: *Libreas - Library Ideas* (23), S. 4–10. Online verfügbar unter <http://libreas.eu/ausgabe23/02voss/> [28.11.2020]
- Vouros, George A.; Panayiotopoulos, Themistoklis; Makedon, Fillia; Ye, Song; Zhang, Sheng; Ford, James et al. (Hg.) (2004): *Data Brokers: Building Collections through Automated Negotiation. Methods and Applications of Artificial Intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Vowinkel, Bernd (2017): Digitale Intelligenz. KI. In: Oliver Stengel, Alexander van Looy und Stephan Wallaschkowski (Hg.): *Digitalzeitalter - Digitalgesellschaft: Das Ende des Industriezeitalters und der Beginn einer neuen Epoche*. Wiesbaden, S. 89–108.
- Wang, Richard Y.; Strong, Diane M. (1996): Beyond Accuracy. What Data Quality Means to Data Consumers. In: *Journal of Management Information Systems* 12 (4), S. 5–33.
- Weizenbaum, Joseph (1977): *Die Macht der Computer und die Ohnmacht der Vernunft*. Frankfurt.
- Werber, Niels (2004): Vom Unterlaufen der Sinne. Digitalisierung als Codierung. In: Alexander Böhnke und Jens Schröter (Hg.): *Analog/Digital - Opposition oder Kontinuum? Zur Theorie und Geschichte einer Unterscheidung*. Bielefeld, S. 81–96.
- Winkler, Hartmut (2004): Medium Computer. Zehn populäre Thesen zum Thema und warum sie möglicherweise falsch sind. In: Lorenz Engell und Britta Neitzel (Hg.): *Das Gesicht der Welt. Medien in der digitalen Kultur*. München, S. 203–213.