



Lothar Kolditz

Big Data - Die große Datenflut, Theorien, Modelle und Berechenbarkeit

Vortrag in der Klasse für Naturwissenschaften und Technikwissenschaften am 9. April 2015

Big Data

Unser Zeitalter ist durch Big Data geprägt, Daten, die laufend gesammelt werden, in großen Mengen dauerhaft im Netz vorhanden und in ständigem Anstieg begriffen sind. Die Daten werden unaufhaltsam geliefert von Mobilfunkgeräten, Notebooks, Tablets usw., von Forschungseinrichtungen und Messinstrumenten im Kosmos und auf der Erde, von Verwaltungen und Betrieben, von den unterschiedlichsten Geräten, die als Computer fungieren. Die ungehinderte Sammlung der Daten ist möglich durch das gewaltige, nicht mehr vorstellbare Anwachsen der Speicherkapazität, die den Petabyte-Bereich (10^{15}) bereits weit überstiegen hat und deren Grenze nicht geortet werden kann.

Es existieren verschiedene Netze nebeneinander, das öffentliche, jedem Teilnehmer zugängliche Internet und das Intranet in Form interner Netze von Firmen, Industriezweigen, Forschungseinrichtungen, Behörden sowie die Netze der Geheimdienste. In den Netzen sind Sachdaten vorhanden, die ein Großteil des Wissens der Menschheit umfassen und persönliche Daten, die die private Atmosphäre praktisch offenlegen, weil auch Verschlüsselung umgangen werden kann. Dieses Wissen im Netz ist zwar nicht fehlerfrei, wird aber immer wieder bearbeitet, um die Fehlerquote möglichst niedrig zu halten.

Alle Bereiche menschlichen Zusammenlebens werden direkt und indirekt berührt, alle können profitieren bei nützlicher Anwendung, aber auch Schaden nehmen bei kriminell oder sorglos und verantwortungslosem Missbrauch. Die Auswertung der Daten wird immer neue Möglichkeiten eröffnen. Entwicklungen und Tendenzen werden prognostiziert. Mit dem Fortschritt entstehen aber auch höhere Anforderungen an ethische Auffassungen und eine wachsende Verantwortung wird notwendig, wenn Rückschläge vermieden werden sollen.

Erzeugen und Auswerten von Daten ist an Datenträger und neuerdings auch an Maschinen gebunden, also an Materialien wie früher die Zeichnungen auf Höhlenwänden, die Keilschrift auf Tontafeln, Abbildungen auf Steintafeln, Darstellungen auf Pergament, dann auf Papier und die Vervielfältigung der Dokumente durch den Buchdruck als eine nur schwache Vorstufe für den heutigen Datenanstieg.

Dabei fällt auf, dass die Dauerhaftigkeit der älteren Datenträger die der neueren übersteigt. Höhlenzeichnungen, Darstellungen auf Ton- und Steintafeln sind haltbarer als Pergament und Papier, die wiederum dauerhafter sind als die heutigen digitalen Datenträger, die der ständigen Pflege und Übertragung auf durch neuere Maschinen lesbare Träger bedürfen. In Sorge um das Verschwinden von Informationen durch Ausfall der neueren Datenträger greifen Initiativen wie Memory of Mankind www.memory-of-mankind.com auf das Fixieren von Wissensextrakten der heutigen Zeit auf alte Materialien wie Tontafeln und ihre sichere Lagerung zurück.

Mit Big Data kommt das Zeitalter der digitalen Revolution voll zur Geltung, die nach über 200 Jahren der industriellen Revolution folgt. Diese Revolution hat schon große Veränderungen eingeleitet. Umfangreiche Lexika, wie die Encyclopaedia Britannica, die ab 1768 herausgegeben wurde, sind als gedruckte Sammlung von Wissen überflüssig geworden, weil sie mit dem ständig steigenden Inhalt des Netzes nicht mithalten können. Ab 2012 erscheint die Encyclopaedia Britannica nur noch digital, und Wikipedia ist eine dazu kaum schlagbare Konkurrenz. Die fortlaufende Vernetzung von Maschinen und Geräten über das Internet bis hin zum Kühlschranks wird weitere Veränderungen mit sich bringen und Geschäftsmodelle beeinflussen.

Wie zu Anfang jeder Revolution gibt es Einschätzungen, die noch nicht abgeklärt sind und vor euphorischer Begeisterung über die neuen Möglichkeiten über das Ziel hinausschießen. In dem Buch „Big Data – das neue Versprechen der Allwissenheit“ [1] wird die Datenflut von verschiedenen Seiten beleuchtet und von mehreren Autorinnen und Autoren behandelt. Chris Anderson [1, S. 124/130] vertritt die Meinung, dass die Datenschwemme das Ende der Theorie herbeiführt und wissenschaftliche Methoden obsolet macht. Er hält „die traditionelle Herangehensweise – Hypothesenbildung, Modell, Test – für veraltet“.

Auch der Softwareunternehmer Stephen Wolfram (A New Kind of Science 2002) [vgl. 2, S.103] kommt angesichts der steigenden Rechenleistung von Computern und der unerschöpflichen Speicherkapazität zu der Schlussfolgerung, dass Computereinsätze an die Stelle von mathematischen Beweisen und Theorien treten werden. Problemlösungen werden mit gewaltigen Rechenleistungen durch Probieren, also empirisch, zu finden sein, so dass Theorien, Beweise und Erklärungen überflüssig werden. Stephen Wolfram hatte festgestellt, dass durch schnelle Computerleistung in zellulären Automaten Musterentwicklungen zu erreichen waren, die sich scheinbar zufällig gebildet hatten. Die Produktion von Datenmustern und damit zusammenhängende Entdeckungen werden mit der sich entwickelnden Computertechnologie nach Wolfram in Zukunft viel schneller erreichbar sein als durch aufwendige Theorien und Beweise vorhersagbar.

Die in diesem Zusammenhang genannten Computereperimente wollen wir als ungeordnet bezeichnen. Sie streben Ergebnisse durch probierendes Vermischen von Daten an.

Diese Auffassungen zur Rolle von Theorien und Modellen im Zeitalter von Big Data und Supercomputern lassen es notwendig erscheinen, Theorien und Modelle in ihrer historischen Entwicklung zu betrachten, ihren genauen Charakter und Inhalt zu bestimmen und daraus ihre richtige Einordnung in die heutige Zeit zu begründen, um Fehlbeurteilungen zu vermeiden.

Die zu betrachtenden Vorgänge sind sehr komplex, so dass es zum besseren Verständnis durchaus angebracht ist, aus ganz verschiedenen Blickwinkeln eine Annäherung zu versuchen, wobei die verwendeten Begriffe wegen ihrer Mehrdeutigkeit zu definieren sind, um fehlerhaften Interpretationen entgegen zu wirken.

Daten, Theorien und Modelle

Theorien enthalten nach der in dieser Ausarbeitung vertretenen Auffassung Regeln oder auch Gesetze, die ein bestimmter für ihr Gebiet zutreffender Datensatz befolgt, wodurch Voraussagen für ablaufende Vorgänge abgeleitet werden können. Mit Theorien sind Datensammlungen zu ordnen.

Aus Theorien folgen Bilder für die Realität, die Modelle genannt und verifiziert oder falsifiziert werden können. Im Ergebnis dieses Prozesses werden Modelle und übergeordnete Theorien korrigiert und der Realität weiter angenähert.

Theorien und Modelle treffen immer nur für einen bestimmten Bereich zu, dessen Grenzen anzugeben sind. Sie sind nicht über alle Grenzen hinaus allgemein gültig.

Betrachten wir dazu konkrete Beispiele und beginnen mit Newton. Modelle, die von Newtons Theorie ausgehen, unterliegen nach Chris Anderson einer veralteten Betrachtung. Bei den Newtonschen Modellen handele es sich nach seiner Ansicht um eher grobe Annäherungen an die Wahrheit, und auf der Ebene der Atome seien sie ohnehin falsch.

Das von Isaac Newton (1643 - 1727) gefundene Gravitationsgesetz lautet:

$$F = G \frac{M_1 \times M_2}{r^2}$$

Die Gravitation F ist eine Kraft, die zwischen den beiden Massen M_1 und M_2 wirkt und mit dem Quadrat des Abstandes beider Massen abklingt. G ist die Gravitationskonstante, die die Dimensionen der Massen und des Abstandes beinhaltet und Dimensionsgleichheit mit der Kraft F herstellt.

Die Massen M_1 und M_2 setzen sich aus Ansammlungen von Atomen und Molekülen zusammen. Der Gültigkeitsbereich des Gravitationsgesetzes betrifft alle Ansammlungen im Makrokosmos. Es gilt jedoch nicht für den Mikrokosmos, also nicht für die Kräfte im Inneren von Atomen und Molekülen.

Diese Grenze hat Chris Anderson in seiner falschen Betrachtung des Charakters von Modellen nicht beachtet.

Streng genommen gilt diese Formel auch nur für den Fall, dass die beiden Massen M_1 und M_2 unbeeinflusst im Raum vorhanden sind, das heißt, dass keine anderen Massen auf sie einwirken. Das ist ein nicht zu realisierender Idealzustand, der aber die Berechnung nach der Formel trotzdem zulässt, da störende Einflüsse anderer Massen in der Regel wegen Geringfügigkeit unberücksichtigt bleiben können. Dennoch ist aber das Ergebnis nur eine Näherung. Die Gravitationswirkung zwischen der Erde und einem Apfel auf dem Baum, die Newton betrachtete, wird von anderen im Kosmos vorhandenen Massen nicht messbar verfälscht. Anders liegen die Verhältnisse bei Ebbe und Flut und dem Einfluss des Mondes und dem der Sonne, letzterer wegen des größeren Abstandes in abgeschwächter Form. Auch bei dem Versuch der genauen Messung der Gravitationskonstanten ist dieser Einfluss von Bedeutung.

Wir betrachten nun komplexere Verhältnisse. Nach dem Ptolemäischen Weltbild steht die Erde im Zentrum. Die Planeten sind an durchsichtigen Sphären befestigt, ganz außen befinden sich die Fixsterne. Dieses Weltbild musste als Modell die Bewegungen der Planeten wiedergeben, so dass auch Vorausberechnungen möglich waren. Den zu beobachtenden rückläufigen Bewegungen von Planeten trug Ptolemäus durch die Annahme Rechnung, dass die Planeten in ihren jeweiligen Sphären Epizyklen bilden, indem sie auf der Umlaufbahn um Punkte kreisförmig rotieren. Auf diese Weise gelangen zutreffende Berechnungen.¹

Es ist also durchaus möglich, mit einem falschen Modell zu richtigen Aussagen zu kommen. Einer tieferen Prüfung hält jedoch ein falsches Modell nicht Stand. Das Ptolemäische musste den Kopernikanischen und Keplerschen Modellen mit der Sonne im Zentrum weichen, wobei Nikolaus Kopernikus (1473 - 1543) kreisförmige und Johannes Kepler (1571-1630) in der Weiterentwicklung elliptische Umlaufbahnen formulierte.

Wie in der Antike setzte sich auch im Mittelalter das heliozentrische Weltbild selbst bei Astronomen nicht sofort durch, obwohl Galileo Galilei (1564 - 1641) mit seinem Teleskop die Venusphasen und die Jupitermonde beobachtet hatte.

Vor allem verwies die Kirche auf die Autorität der Bibel. Josua sprach bei dem Sieg Israels über die Amoriterkönige, als er für die Fortführung der Schlacht eine Verlängerung des Tages wünschte: „Sonne, stehe still im Tale Gibeon“ (Josua 10, Vers 12). Ihre Wirkung zeigte sich bei Tycho de Brahe (1546 – 1601), der ein gemischt geo- und heliozentrisches Modell formulierte, bei dem sich die Erde im Mittelpunkt befindet und vom Mond und von der Sonne mit den Planeten umrundet wird [3].

Analoge Feststellungen ergeben sich bei der Entwicklung der Atommodelle, die für die Wirkungen der Kräfte im Atominneren gelten, nur ist hier der zeitliche Ablauf wesentlich schneller.

Das Atommodell von Niels Bohr (1885 – 1962), das Quanteneffekte berücksichtigte, weist den Elektronen um den Kern bestimmte Bahnen zu. Es gestattet die Deutung des Wasserstoffspektrums in den Grundzügen. Die Feinstruktur des Spektrums wurde nicht wiedergegeben, so dass Arnold Sommerfeld (1868 – 1951) an Stelle der Kreisbahnen elliptische Umläufe annahm. Aber auch dieses Modell hielt tieferen Überprüfungen nicht Stand, was die Aufstellung des wellenmechanischen Modells verursachte mit Einführung von Wahrscheinlichkeitsdichten für die Aufenthaltsräume der Elektronen um die Atomkerne. Die Schrödinger-Gleichung wurde von Erwin Schrödinger (1887 – 1961) aufgestellt, in der die Wellenfunktion ψ enthalten ist, deren Quadrat die Dichte der Aufenthaltswahrscheinlichkeit der Elektronen im Atom angibt. Sie gestattet, chemische Bindungen der Atome zu berechnen. Exakt lösbar ist sie allerdings nur für das Wasserstoffatom. Für kompliziertere Fälle müssen numerische Näherungsverfahren zur Ergebnisfindung angewendet werden.

In der Physik erklärt das Standardmodell der Elementarteilchen die Struktur der Materieteilchen in den Atomkernen, stellt die kosmische Entwicklung nach dem Urknall in heutiger Auffassung dar und bezieht die elektromagnetische Kraft sowie die starke und die schwache Kernkraft ein, nicht

¹ Claudius Ptolemäus lebte um 100 – 160 in Alexandria. Aristarchos von Samos (310 v.Chr. – 230 v.Chr.) vertrat schon über 300 Jahre vorher das richtige heliozentrische Modell. Es wurde aber nicht durchgehend anerkannt.

jedoch die Gravitation. Es wird versucht, das Standardmodell zu verbessern, aber auch davon abweichende neue Modellformulierungen wie die Stringtheorie und die Schleifenquantengravitation werden erprobt. Wir sind also keineswegs bei einem endgültigen Modell angelangt.

Aus diesen Betrachtungen lässt sich Folgendes ableiten:

Mit Theorien werden Daten strukturiert, zusammengefasst und geordnet. Eine große Datensammlung erfordert also zur Bearbeitung und Erschließung geradezu die Theorien und macht sie keinesfalls überflüssig [s.a. 2, S. 104].

Theoretische Zusammenhänge gelten für bestimmte Bereiche. Bei Grenzüberschreitungen verlieren sie ihre Anwendungsmöglichkeit, wie das Beispiel Gravitation und Wechselwirkungen im Atomkernbereich lehrt.

Modelle sind Näherungsannahmen, deren Gültigkeit an die richtige Vorhersage von Vorgängen im System geknüpft ist und die mit den Theorien verbunden sind. Falsche Annahmen führen bei Tiefenprüfung zur Ablösung des Modells und zur Veränderung der Theorie.

Mit der Näherung eines Modells an reale Bedingungen erhöhen sich in der Regel die Rechenanforderungen zur Ergebnisfindung. Modelle sind wie die Theorien in bestimmten Bereichen gültig, also keineswegs allgemein zutreffend.

Datenauswertung

Im Datennetzwerk eingespeicherte Daten können trotz ständigen Anstiegs durch Suchfunktionen wiedergefunden werden, die sich immer in Verbesserung befinden. Die Arbeitsvorschrift für den Computer ist in Software enthalten und durch Algorithmen gegeben. Welche Wirkungen sich aus digitalen Textvergleichen mit Hilfe geeigneter Algorithmen ergeben, lehrt das Aufdecken von Plagiaten bei eingereichten akademischen Abschlussarbeiten.

Die Auswertung spezieller Daten, z. B. Forschungsdaten, erfolgt über entsprechend angepasste Software.

Negative Auswirkungen des Datenflusses durch absichtlich herbeigeführte Überlastung eines Netzteils müssen mit geeigneter Software bekämpft werden, wie auch immer wieder Gegenmaßnahmen erforderlich sind bei krimineller Nutzung der Netzmöglichkeiten.

Im Prinzip ist es für einen Teilnehmer möglich, der Zugang zum Datennetz hat, mit einer mächtigen Suchfunktion jeder Frage nachzugehen und den jeweiligen Wissensstand zu erfassen. Für den Einzelnen ist es jedoch nicht notwendig, sämtliche Zusammenhänge zu erkunden, da bereits geleistete Vorarbeit genutzt und Zusammenfassungen von Daten verwendet werden können, wie sie im Netz vorhanden sind.

Für Geheimdienste sind vor allem persönliche Daten von Interesse. Verschlüsselung und Dechiffrierung stehen in ständigem Wettbewerb. Ein absolutes Geheimnis gibt es nicht, weil jeder Verschlüsselungsmaßnahme logische Regeln zu Grunde liegen, die durch logische Decodierung im Prinzip aufgedeckt werden können. Es wird prognostiziert, dass der Quantencomputer, dessen Grundlagen angedacht sind, der aber noch nicht realisiert ist, eine Entschlüsselung unmöglich macht, weil in den dabei verwendeten Qubits mit verschränkten Teilchen gearbeitet werden kann. Das ist eine Vermutung, die als solche zunächst besteht, der Beweis für die Richtigkeit ist damit noch nicht erbracht.

Die Auswertung von Daten wird besonders durch Geheimdienste ständig ausgebaut, einen kleinen Einblick dazu ermöglichten die Enthüllungen Edward Snowdens. Aber auch zur Anbringung gezielter Werbung erfolgen Datenauswertungen durch große Datensammler wie Google oder Amazon. Die Auswertung erfordert immer entsprechend entworfene Algorithmen.

Jeder Teilnehmer, der das Netz benutzt, um Wissen zu erfragen, der online Einkauf tätigt, Nachrichten vermittelt oder empfängt, erzeugt Daten, die ihn charakterisieren und die trotz Datenschutzbemühungen von Dritten zu ökonomischen und weiteren Zwecken verwendet werden.

Diese Felder sind beliebig erweiterbar mit Fortschreiten der Computerleistung und dem Ausbau der Algorithmen. So kann für die Polizei die Häufung von Verbrechen in bestimmten Gebieten zu bestimmten Zeiten Anlass geben, geeignete Maßnahmen zu ergreifen, um noch nicht begangene Verbrechen einzuschränken oder zu verhindern.

Der Erfolg der Datenauswertung wird durch die Güte des Algorithmenentwurfs bestimmt, der Datenkorrelationen durch den Computer veranlasst. So ist der Vergleich von Anruflisten für die Polizei ein wichtiger Hinweis auf das reale Geschehen. In vielen Fällen ist dabei der Inhalt der Gespräche nicht wesentlich. Es kommt auf die Kopplung dieser Daten mit anderen an und deren Auswertung.

Geordnete Computereperimente

Eine gezielte Datenauswertung zu wissenschaftlichen Untersuchungen ist nur mit der Unterstützung von Theorien und Modellen erfolgreich. Ungeordnete oder empirische Computereperimente (vgl. S. 2) führen zwar zu Mustern als Phänomen, aber nicht zur Kenntnis der Grundlagen, die die Musterbildung veranlassen. Zur grundlegenden Auswertung von wissenschaftlichen Daten bedarf es der Anweisungen an den Computer durch entsprechende Algorithmen.

In einem Buch „Die Berechnung der Welt – von der Weltformel zu Big Data“ geht Klaus Mainzer [2] auf die empirischen Möglichkeiten von Big Data in Verbindung mit Superrechnern ein [2, S. 24, 25], analysiert sie, vertritt aber nicht die Auffassung von Stephen Wolfram und Chris Anderson, sondern gibt im Gegenteil mit seinem Buch „ein Plädoyer für die Besinnung auf die Grundlagen, Theorien, Gesetze und Geschichte, die zu der Welt führen, in der wir heute leben“ [2, S. 14].

Gabriele Gramelsberger hat zum Thema Computereperimente ein Buch geschrieben [4], das analytisch die historische Entwicklung des Rechnens im Hinblick auf den heutigen Computereinsatz sehr gut beschreibt.

Die in diesem Abschnitt behandelten Computereperimente werden hier zum Unterschied von den auf Seite 2 erwähnten ungeordneten Experimenten als geordnete Computereperimente bezeichnet.

Unter geordneten Computereperimenten wird die Verwertung von großen Datenmengen in einer durch Theorien, Gesetze und Modelle vorgezeichneten Ordnung verstanden, die in den verwendeten Algorithmen enthalten ist. Diese Computereperimente unterscheiden sich also grundsätzlich von den Computereperimenten Wolframs, die einen probierenden Einsatz großer ungeordneter Datenmengen vorsehen.

Komplizierte Vorgänge in der Natur sind nicht linear. Die dafür zu formulierenden mathematischen Differentialgleichungen sind in vielen Fällen nicht algebraisch direkt lösbar und können nur numerisch angenähert werden.

Ein Beispiel dafür ist die Strömungsdynamik. Die im 19. Jahrhundert im Wesentlichen von Claude Louis Marie Henri Navier (1785 -1836) und George Gabriel Stokes (1819 -1903) entwickelten Bewegungsgleichungen für Strömungen in Flüssigkeiten und Gasen [4, S. 71] gehören zu den partiellen Differentialgleichungen. Sie sind für reale Flüssigkeiten nicht direkt lösbar, sondern nur durch numerische Simulation für die Praxis erschließbar, wozu die heutigen Computer mit ihren Rechengeschwindigkeiten gute experimentelle Bedingungen bieten.

Die Näherungsverfahren zur numerischen Annäherung an die Lösung von Differentialgleichungen sind unterschiedlicher Art. John von Neumann (1903 – 1957) verwendete die Differenzenmethode zur Berechnung von Differentialgleichungen auf Computern, die von Joseph-Louis Lagrange (1736 – 1813) bereits 1759 entwickelt wurde [4, S. 76]. Es wird der Übergang von Differenzen Δy und Δx zu Differentialen dy und dx simuliert. Die damaligen Computer (1944) waren aber noch zu wenig leistungsfähig. Mit heutigen Computern ergeben sich ganz andere Möglichkeiten. In jedem Fall bleibt es aber bei der Näherungslösung, wenn es auch gelingt, relativ nahe an das reale Verhalten zu gelangen.

Die Näherungsverfahren werden den jeweils zu bearbeitenden Systemen angepasst. Für quantenchemische Berechnungen unter Zugrundelegung der Schrödinger-Gleichung gelten z. B. andere Verfahren als die von John von Neumann verwendeten. Er war führend in der frühen Computerentwicklung, weil er die Bedeutung der Computer für numerische Näherungslösungen erkannte.

Die heutigen Fortschritte in der Wetterprognose sind mit Computersimulationen der Strömungsverhältnisse in der Atmosphäre und ihrer Kopplung mit der Hydrosphäre verbunden. Ihre Grundlage sind die Ergebnisse weltweiter Messpunkte [4, S. 105/139]. An diesem Beispiel wird die Differenz zu der irrtümlichen Meinung deutlich, dass durch Big Data nunmehr Theorien und Modelle überflüssig

werden. Daten ohne die ordnende Theorie der zugrunde liegenden Navier-Stokes-Gleichungen ergeben noch keine Information. Für die Auswertung sind entsprechende Algorithmen notwendig.

Ein weiteres Beispiel für geordnete Computerexperimente stammt aus den Biowissenschaften. Zur Aufklärung molekularbiologischer Prozesse wurden elektronenmikroskopische Aufnahmen von biologischem Material und bei kristallinem Material auch Röntgenbeugungsspektren hergestellt. Sie sind Momentaufnahmen der räumlichen Positionen der Atome während eines molekularbiologischen Vorganges. Die Daten von in Abständen von Nanosekunden erhaltenen Aufnahmen wurden dann von Superrechnern verglichen, um gewissermaßen aus den Momentaufnahmen Videodarstellungen zu erhalten und so die ablaufenden Vorgänge zu ergründen. Dadurch wurden Hinweise für medikamentöse Angriffspunkte bei HIV-Infektionen gefunden, weitere Aufklärungen zum Ablauf der Photosynthese erhalten und die Zerlegung von Kohlenhydratpolymeren in monomere Zucker verfolgt. Es handelt sich also bei diesen Untersuchungen um Experiment gestützte Datenverwendung auf Basis von Theorien und nicht um bloße Nutzung einer hohen Anzahl von Daten [5].

Bei einem Riesenanstieg von Daten etwa bei kosmischen Experimenten ist eine sofortige vollständige Auswertung nicht durchführbar. Daten werden abgespeichert, die Auswertung erfolgt zeitverzögert.

Im Falle der Kollisionsexperimente des Large Hadron Collider im CERN fallen so viele Daten an, dass durch eine komplizierte algorithmisch gesteuerte Datenauswahl eine Aussonderung von Daten erfolgt, so dass nur ein Bruchteil zur direkten Auswertung gelangt, an der sich neben dem Rechenzentrum des CERN noch z. Zt. über 170 Rechenzentren in 40 Ländern beteiligen [6].

Geordnete Computerexperimente sind nicht auf naturwissenschaftliche Bereiche beschränkt, sondern haben längst auch das Feld der Sozial- und Geisteswissenschaften besetzt. Auch hier zeigt sich das Bild der informatorischen Computerunterstützung für Verfahren, die in der Vorcomputerzeit bereits bekannt und angedacht waren. Linguistische Analysen, Stilvergleiche, Untersuchungen zur Autorenurheberschaft, Vergleiche und Zusammenführung von verstreuten archäologischen Funden bieten sich für eine Computerlösung an in wirksamer Ergänzung erprobter hermeneutischer Verfahren, z. B. linguistische Analysen der Paulusbriefe, von Shakespearetexten, Aufzeichnungen Thomas von Aquins; vgl. Gerhard Lauer „Die digitale Vermessung der Kultur – Geisteswissenschaften als Digital Humanities“ [1, S. 99/116],

Wiederum zeigt sich hier die Verwendung von Messwerten, nämlich der Textgrundlagen oder der digital erfassten Fundstücke verbunden mit der experimentellen Frage der Datenverwertung, die durch Algorithmen ausgedrückt wird, in denen die Theorie steckt.

Besonders bei soziologischen Analysen muss sorgfältig darauf geachtet werden, dass durch die Datenauswahl nicht das Ergebnis verzerrt wird. An Hand von Studien lässt sich leicht verfolgen, dass sich die Datenauswahl nach der Tendenz des Geldgebers richten kann, was dann die Objektivität in Zweifel zieht.

Für alle Materialanalysen gilt, dass nicht allein die Güte der Instrumente und ihr richtiger Einsatz das Ergebnis bestimmt, sondern vor allem die Materialauswahl in der verwendeten Durchschnittsprobe. Für soziologische Untersuchungen betrifft dies die Art der Datenermittlung und die Datenauswahl.

Berechenbarkeit

Angesichts der hohen Rechenkapazität der Supercomputer oder von im Netzwerk zusammengeschlossenen Computern ergibt sich die Frage, ob Grenzen der Berechenbarkeit existieren. Diese Betrachtung soll aber nicht ausschließlich nach den Regeln der Berechenbarkeitstheorie der theoretischen Informatik erfolgen, die als berechenbar diejenigen Probleme bezeichnet, die mit einer Maschine oder dem mathematischen Modell einer Maschine, z. B. einer Turing-Maschine lösbar sind und denen ein terminierter Algorithmus zu Grunde liegt. Es soll vielmehr von Maschinen abstrahiert werden.

Wir geraten damit auch an eine Kategorie von Annahmen, die nach den Unvollständigkeitssätzen von Kurt Gödel (1906 – 1978) weder beweisbar noch widerlegbar sind. In vereinfachter Form lauten die zusammengefassten Unvollständigkeitssätze: Jedes formale System, welches widerspruchsfrei ist

und elementare Arithmetik erlaubt, ist unvollständig bezüglich der Aussage dieser Arithmetik, was auch bedeutet, dass es in Teilbereichen weder bewiesen noch widerlegt werden kann. Es ist also in dieser Hinsicht nicht entscheidbar.

Alle Abläufe in der Natur sind komplex und in der Regel nicht linear, d. h. die wirksamen Funktionsabläufe sind nicht durch gerade Linien, sondern durch mehr oder weniger komplizierte Kurven erfassbar. Linearität tritt nur scheinbar auf wie bei einer Feder, deren Anfangsdehnung bei Anbringung von Gewichten näherungsweise als linear angenommen werden kann, bei stärkerer Belastung erfolgt jedoch deutliche Abweichung von der Linearität.

Alle diese nicht linearen komplexen Vorgänge unterliegen den gesetzlichen Bedingungen eines deterministischen Chaos. Das ist kein Chaos im landläufigen Sinne mit einem völligen Durcheinander, sondern durchaus ein wohlgeordneter Vorgang, bei dem das Zusammenwirken nichtlinearer Abläufe im Einzelnen durch Naturgesetze gesteuert wird. Das gesamte System ist den Naturgesetzen unterworfen, es zeigt aber keine konstante Periodizität. Es sind nur Wahrscheinlichkeitsaussagen möglich, wie das auch von der Wettervorhersage bekannt ist, vgl. auch [7,8].

Die Wahrscheinlichkeitsaussage für ein Ergebnis heißt aber, dass der betreffende Wert nicht exakt eintreten muss. Es kann sich auch um ein anderes Ergebnis handeln, das in einem gewissen Bereich daneben liegt. Je weiter die Zeitvorschau gehen soll, umso breiter wird auch der Bereich, in dessen Grenzen das Ergebnis liegt. Es gibt für jeden Ablauf des deterministischen Chaos eine Zeit, über die hinaus eine Vorhersage nicht mehr möglich ist, die sogenannte Ljapunow-Zeit.²

Die Ljapunow-Zeit beruht auf dem Konzept des Ljapunow-Exponenten, der ein Maß für das durchschnittliche Wachstum einer infinitesimalen Abweichung in den Ausgangsbedingungen von Systemen darstellt.

Zur Erläuterung kehren wir noch einmal zum Kopernikanisch-Keplerschen System der Planetenbewegung zurück. Es gilt das Gravitationsgesetz, wobei nunmehr mehr als zwei Massen Berücksichtigung finden müssen. Nur das Zweikörperproblem ist exakt lösbar. Alle höheren Probleme erfordern iterative Lösungen. Das Sonnensystem ist ein determiniert chaotisches System, Die große Masse der Sonne gegenüber den Planetenmassen erlaubt eine Näherungslösung für die Bewegungen der Planeten. Eine computergestützte numerische Integration der Evolution des Sonnensystems von etwa 100 Millionen Jahren führte zur Ermittlung einer Ljapunow-Zeit von 4 Millionen Jahren [9]. Innerhalb dieser Zeit sind Rechnungen von Standorten und Ereignissen möglich sowohl für die Zukunft als auch für die Vergangenheit, nicht aber darüber hinaus. Ergebnisse außerhalb dieser Zeit sind mit großer Unsicherheit behaftet, geringe Einflüsse auf das System würden sich merkbar auswirken. Die Berechenbarkeit ist dann nicht mehr gegeben, das Wahrscheinlichkeitsintervall wird zu groß.

Das ganze Universum unterliegt den Bedingungen des deterministischen Chaos. Für alle Systeme gelten Ljapunow-Zeiten, die von der spezifischen Dynamik der jeweiligen Bereiche abhängen. Sie bestimmen die Grenzen der Berechenbarkeit.

Die Weltformel

Es ist nützlich, zunächst Klarheit darüber zu schaffen, was unter Weltformel verstanden wird. In der höchsten Kategorie wird die Weltformel als Theory of Everything (TOE) aufgefasst, mit der alle Vorgänge im Kosmos ausgehend von einem Urgesetz erfasst und erklärt werden. Dieses Kriterium der allgemeinen Weltformel ist höher angebunden als das Bestreben der Einbeziehung der Gravitationskraft in die Quantenphysik und in die Vereinheitlichung der drei Grundkräfte elektromagnetische Kraft, starke und schwache Kernwechselwirkung (s. auch S. 4). Albert Einstein (1879 – 1955) und Werner Heisenberg (1901 – 1976) beschäftigten sich ohne Erfolg damit, und auch bis heute ist noch keine Lösung dieses Problems gelungen. Stephen Hawking (geb. 1942) hatte sich zunächst optimis-

² Alexander Michailowitsch Ljapunow (1857-1918) war ein russischer Mathematiker, Mitglied der Russischen Akademie der Wissenschaften, von dem bedeutende Beiträge auf den Gebieten der Differentialgleichungen, der Stabilität von Systemen, der Potenzialtheorie und der Wahrscheinlichkeitstheorie stammen. Er lehrte ab 1885 an der Universität und Technischen Hochschule in Charkow und ab 1901 an der Petersburger Universität, wo er als Ordentlicher Professor den Lehrstuhl für Angewandte Mathematik innehatte.

tisch zu einer allgemeinen Fassung der Weltformel bekannt, schloss aber dann eine positive Lösung aus nach Einbeziehung von Gödels Unvollständigkeitssatz.

In den Ausführungen zu „Kollektivität und Emergenz – die Weltformel“ [7] wurde zusammenfassend festgestellt, dass die wissenschaftlichen Befunde auf den hierarchischen Aufbau des Kosmos aus Teilbereichen hinweisen, bei denen emergente Leistungen aus kollektivem Zusammenwirken innerhalb der Bereiche hervorgehen. So kann einem einzelnen Gasmolekül kein Druck zugeordnet werden. Der Gasdruck ist ein kollektives Ergebnis der Impulswirkung von Gasteilchen auf die Gefäßwände.

In den Teilbereichen des Kosmos sind reduktive Schlüsse möglich, also Schlüsse aus den Eigenschaften der kleinsten Teilchen auf das Ganze, nicht jedoch über die Grenzen der Teilbereiche hinaus, wo jeweils andere Bedingungen gelten. Eine umfassende reduktive Ableitung ist damit unmöglich und die Aufstellung der Weltformel aus dieser Sicht nicht erreichbar.

Das Weltformelproblem soll nun auch aus Sicht der Datenflut in Verbindung mit Modellen beleuchtet werden. Es war klar geworden, dass mit der Erhöhung des Anspruchs auf die Modellleistung eine Verbesserung der Modelle einher zu gehen hat, mit der in der Regel erhöhte Anforderungen an die Rechentechnik verbunden sind. Ein Modell ist immer eine Annäherung an die Realität und nicht ihr reales Abbild. Vereinfachende Annahmen im Modell sind Näherungen und dienen der Reduzierung der erforderlichen Rechentechnik. Mit immer weiter verbesserten Modellen nähern wir uns den realen Verhältnissen, erreichen sie aber nicht. Es handelt sich um eine asymptotische Annäherung, die erst im Unendlichen mit dem vollständigen Wissen übereinstimmt.

Ein Laplacescher Dämon, der alles Wissen in sich vereint, wäre also im Endlichen nicht erreichbar.

Mit dem Wissen, das im Netz vorhanden ist, verhält es sich ähnlich. Eine gewisse Fehlerquote wird immer vorhanden sein, an deren Ausbesserung stets gearbeitet wird. Vollständigkeit ist jedoch nicht erreichbar analog zu dem Modellbeispiel.

Damit folgt auch aus dieser Betrachtung der Schluss, der zur Unerreichbarkeit einer Theorie von Allem oder einer allgemeinen Weltformel führt. Die Steigerung der Datenflut bringt im Endlichen keine Lösung mit sich.

Literatur

- [1] Big Data – Das neue Versprechen der Allwissenheit, Redaktion: Heinrich Geiselberger und Tobias Moorstedt, Suhrkamp Verlag Berlin 2013, 309 S.
- [2] Klaus Mainzer, Die Berechnung der Welt – Von der Weltformel zu Big Data, Verlag C.H.Beck, München 2014, 352 S.
- [3] Dennis Danielson und Christopher M. Graney, Kopernikus, Revolution mit Hindernissen, Spektrum der Wissenschaften, Heft 10 [2014], S. 54-59.
- [4] Gabriele Gramelsberger, Computereperimente – Zum Wandel der Wissenschaft im Zeitalter des Computers, transcript Verlag Bielefeld, 2010, 313 S.
- [5] Till Rudack, Juan Perilla u. Klaus Schulten, Die Geheimnisse des Lebens berechnen, Spektrum der Wissenschaften, Heft 11 [2014] S.86-95.
- [6] Andreas Stiller, Licht ins Dunkel – CERN startet Run 2 des Large Hadron Collider, c't magazin für computertechnik Heft 6, 2015 S. 74-77.
- [7] Lothar Kolditz, Kollektivität und Emergenz – Die Weltformel, Sitzungsberichte der Leibniz-Sozietät der Wissenschaften, Bd. 105 [2010] S.91-106..
- [8] Lothar Kolditz, Deterministisches Chaos und Gesellschaft, Sitzungsberichte der Leibniz-Sozietät der Wissenschaften, Bd. 110 [2011] S. 107-122
- [9] G. Sussman, J. Wisdom, Chaotic Evolution of the Solar System, Science 257 [1992] 56-62.