

Gerhard Lauer

Die digitale Vermessung der Kultur

Geisteswissenschaften als Digital Humanities

Die Last der Revolutionen ergibt sich aus dem Umstand, dass sie das Leben verändern können. Die Digitalisierung ist eine solche Revolution. Sie verändert ziemlich viele Lebensbereiche, das auch noch sehr schnell und ohne dass ein Ende abzusehen wäre. Sie verändert auch die Geisteswissenschaften. Mit dem Label »Digital Humanities« wird dieser Wandel ausgeflaggt, aber wie in vielen ähnlichen Fällen ist auch dieser Begriff eher ungenau. Die Unschärfe der Bezeichnung hat zunächst damit zu tun, dass man ab dem Aufkommen des Computers in den sechziger Jahren üblicherweise von »Humanities Computing« sprach; erst seit der Jahrtausendwende setzte sich dann im Zusammenhang mit der Verbreitung des Internets der Begriff »Digital Humanities« zur Beschreibung der computergestützten Geisteswissenschaften durch. Hießen entsprechende Publikationen in den achtziger Jahren noch *The Humanities Computing Yearbook*, kamen sie nach 2000 mit Titeln wie *A Companion to Digital Humanities* auf den Markt. Mit dem Namenswechsel wird auch die Klammer größer, sie umfasst nun nicht länger allein textwissenschaftliche Fächer, sondern auch Disziplinen wie die Geschichte oder die Archäologie, die Musik- und die Kunstwissenschaften. Die Verbände benennen sich ebenfalls um. Statt »Association for Literary and Linguistic Computing« heißt der europäische Verband nun »The European Association for Digital Humanities«; die Fachzeitschriften tun es den Verbänden nach. Die digitale Revolution kommt also auch in diesen Fächern an. Kenner der Fachgeschichte wie beispielsweise Willard McCarty oder Matthew Kirschenbaum führen die

wachsende Akzeptanz auf diese Erweiterung zurück: Die Digital Humanities haben die Nische verlassen und sind sichtbar geworden unter den Geisteswissenschaften (Kirschenbaum 2010).

Der Namenswechsel markiert nicht einfach den nächsten »turn«, man reagiert damit vielmehr auf ein Problem, das die Wertehierarchie der kulturwissenschaftlichen Disziplinen betrifft. Das Rechnen gehörte lange Zeit so gut wie nicht zum Methodeninventar der Geisteswissenschaften (vgl. dazu auch den Beitrag von danah boyd und Kate Crawford in diesem Band), wobei methodische Innovationen in diesem Bereich ohnehin eine eher untergeordnete Rolle spielen. In den Natur- und Lebenswissenschaften ist der Einsatz von Computern längst selbstverständlich; an spezialisierte Disziplinen wie Computational Physics, Bio-, Geo- und andere Bindestrich-Informatiken hat man sich gewöhnt; die Vergabe des Physiknobelpreises an Gerd Binnig und Heinrich Rohrer, die 1986 für die Entwicklung des Rastertunnelmikroskops ausgezeichnet wurden, belegt den Stellenwert methodischer Innovationen in diesen Fächern.

In den Geisteswissenschaften ist das anders. Ihre historisch-hermeneutischen Verfahren werden auf die hellenistische Textkritik oder den Humanismus, auf die Altphilologie eines Christian Gottlob Heyne oder Friedrich August Wolf sowie Philosophen des 19. Jahrhunderts zurückgeführt, man denke etwa an Wilhelm Dilthey. Die Tradition adelt die humanistischen Disziplinen, und wenn sich doch einmal der Wunsch nach Modernität und Innovation regt, wird dieser durch die theoretischen »turns« hinreichend bedient. So weiß man sich stets à jour, und kulturkritische Urteilsroutinen über den Gegensatz von kalter, technischer Zivilisation und warmer, verstehender Kultur tragen das ihre dazu bei, dass Geisteswissenschaften Innovation in ihren Methoden und in ihrer Methodik selten belohnen, zumal wenn sie mit dem Computer und dem Internet verknüpft sind.

Eine rechnende Geisteswissenschaft muss vor diesem Hintergrund geradezu als ein Oxymoron erscheinen. »Digital Humanities« ist da ein handlicheres Label für nicht ohne Weiteres zu integrierende Erweiterungen des Methodeninventars, die über qualitativ-interpretierende Verfahren hinausgehen. Die Unschärfe der Bezeichnung ist dabei den epistemischen Passungsschwierigkeiten von Methodik und Disziplin geschuldet. Und doch ist genau das gemeint: die methodische Ergänzung der Geisteswissenschaften um im weitesten Sinne rechnende Verfahren, für die der Computer und das Internet nützlich sind (Lunenfeld et al. 2012).

Etablierte Problemstellungen anders anzugehen und auch neue, bislang kaum zu beantwortende Fragen zu bearbeiten – das macht die Digital Humanities aus, nicht ein weiterer »(computational) turn«. Gelegentlich gibt sich diese Revolution radikal, etwa mit dem »Digital Humanities Manifesto«. Dominiert jedoch wird die Diskussion, die in Foren wie »Humanist«, »Digital Humanities im deutschsprachigen Raum« oder beispielsweise in den Blogs von Dan Cohen und Martin Mueller geführt wird, von einer geradezu topischen Bescheidenheitsrhetorik, die an das philologische Ethos des Dienstes am Text anknüpft.

I.

Diese Selbstbescheidung hat vor allem mit der Herkunft der Digital Humanities aus den Traditionen der Philologie zu tun. Philologie beginnt vor dem Interpretieren, nämlich mit dem Sammeln der Daten (seien es Texte oder Dinge), dem Katalogisieren und Auszeichnen, dem Vergleichen und Inbeziehungsetzen, dem Ordnen überhaupt. Diese elementaren Methoden oder »scholarly primitives« sind es, die die Einheit der Fächergruppe ausmachen (Unsworth 2000). Ein solches Elementar-

unternehmen war schon die von Theodor Mommsen angeregte Sammlung aller Inschriften des Römischen Reiches, geordnet nach den jeweiligen Provinzen. Für dieses 1853 begonnene monumentale Projekt hatte Mommsen vorab keine Frage vorgegeben, er wollte die Daten zukünftigen Studien zur Verfügung stellen. Heute umfasst das »Corpus Inscriptionum Latinarum« mehr als 180 000 Inschriften. Es bringt zum Ausdruck, was den wissenschaftsgeschichtlichen Übergang zur Neuzeit kennzeichnet: der neuartige Umgang mit Daten. Wissenschaftshistorikerinnen wie Lorraine Daston haben gezeigt (Daston/Park 2002), wie in den aufkommenden modernen Wissenschaften an die Stelle der Exotika und Rarissima der Kunstkammern und Jahrmärkte die kontrollierte Genauigkeit der Beobachtung beim Sammeln der Daten trat. Das, und nicht die Hermeneutik, macht Wissenschaften modern, auch die Philologien. Und diese Traditionslinie schreiben nun auch die Digital Humanities fort.

Mommsen war keineswegs der Einzige, der vor alle historisch-hermeneutische Aneignung das Sammeln von Daten gesetzt hat. Etwa zur selben Zeit entwickelte der englische Mathematiker Augustus de Morgan die Idee, durch das Auszählen der Silben Stilunterschiede zu identifizieren. Als beispielhafte Anwendung brachte er die Frage nach der Urheberschaft der Briefe ins Spiel, die dem Apostel Paulus zugeschrieben werden. Um zu entscheiden, ob alle diese Briefe tatsächlich von Paulus verfasst wurden, setzte de Morgan nicht auf Hermeneutik im engeren Sinne, sondern er ermittelte die durchschnittliche Wortlänge in den einzelnen Schreiben. Seine Hypothese war einfach: In den Briefen eines Autors sollten die Wortlängen immer ähnlicher verteilt sein als in Texten unterschiedlicher Verfasser. Wenige Jahre später versuchte Thomas Corwin Mendenhall, ein amerikanischer Meteorologe und leidenschaftlicher Shakespeare-Leser, den Stil des englischen Schriftstellers genauer zu

charakterisieren, indem er nicht nur Wortlängen, sondern auch die relative Häufigkeit der Worte in gleichmäßig segmentierten Textabschnitten bestimmte. Seine Analysen ergaben charakteristische Kurven, die es ihm ermöglichten, den individuellen Stil Shakespeares von dem von Zeitgenossen wie Francis Bacon zu unterscheiden. Sein Verfahren hat Mendenhall später auch an Autoren wie Dickens, Thackeray oder Mill erprobt. Unabhängig von de Morgan und Mendenhall unternahm der polnische Altphilologe und Philosoph Wincenty Lutosławski Ende des 19. Jahrhunderts den Versuch, die umstrittene Chronologie der Schriften Platons zu klären, indem er verschiedene formale Texteigenschaften oder »Stylème« auszählte. Lutosławski ging dabei von der Hypothese aus, dass sich formale Eigenschaften kontinuierlich entwickeln, so dass in zeitlicher Nähe entstandene Texte größere Ähnlichkeiten aufweisen sollten. Allgemeiner nahm Lutosławski sogar an, dass man stilometrische Muster identifizieren könne, anhand deren Schriftsteller sich besser würden unterscheiden lassen als mithilfe hermeneutischer Verfahren (vgl. Grzybek 2007).

Solche und ähnliche, um weitere Stileigentümlichkeiten ergänzte Ansätze gibt es nicht nur in der philologischen Tradition des 19. Jahrhunderts, in der russischen Literaturwissenschaft, im französischen Strukturalismus oder in der quantitativen Linguistik. Diese Verfahren spielen heute eine Rolle, wenn es beispielsweise um Autorschaftsattributions geht, und sie haben längst auch Einzug in die Forensik, in den investigativen Journalismus oder die Arbeit der Geheimdienste gehalten (Joula 2008).

Selbst wenn die Philologie in solchen Versuchen eine allzu enge Beziehung mit der Mathematik eingegangen sein mag, finden sich ähnliche Ansätze auf einem Feld, das man zum traditionellen Kerngeschäft der Philologie zählen kann: die Edition. Es war im Jahr 1949, als nur wenige Literaturwissenschaftler

wussten, dass es so etwas wie Computer überhaupt gab, als der Jesuitenpater Roberto Busa im Zuge seiner Studien zum Begriff der Präsenz im Werk des heiligen Thomas auf den Gedanken kam, dass Computer ein nützliches Instrument sein könnten, um den wechselnden Begriffsgebrauch in diesem kaum zu überblickenden Werk zu erfassen, also um beispielsweise Kookkurrenzen (das gemeinsame Auftreten zweier Wörter) zu ermitteln, Parallelstellen aufzuzeigen und so eine Konkordanz zu erstellen. Die Forschungsfrage und die elementaren methodischen Verfahren waren nicht neu. Dafür Computer zu verwenden, war allerdings revolutionär. Pater Busa gelang es, Thomas J. Watson, den Gründer von IBM, für seine Pionieridee zu begeistern. So entstand die erste digitale Edition auf Lochkarten, das »Corpus Thomisticum«, das ein kulturgeschichtlich bedeutendes Werk auf eine vollkommen neue Weise aufschlüsselte und das heute sowohl in 56 Bänden auf Papier als auch im Netz zugänglich ist. In den sechziger Jahren folgte eine erste elektronische Fassung der *Modern Language Association International Bibliography* (MLAIB), einer periodischen Fachbibliographie aller modernen Philologien, die mit Telefonkopplern durchsucht werden konnte. Dann setzte die retrospektive Digitalisierung des kulturellen Erbes ein, die immer mehr Werke und Lexika wie etwa das *Deutsche Wörterbuch* der Brüder Grimm, historische Wörterbücher wie den *Krönitz* oder Regionalwörterbücher umfasste. Wer den Gebrauchsweisen von Begriffen nachgehen will, gleich ob es um Wörter wie »Seele« oder »Gehirn«, »Türken« oder »Zeus«, »Biermolke« oder »Glasflügler« geht, kann im Internet unter Adressen wie woerterbuchnetz.de in einer Fülle von Wörterbüchern detaillierte Belegstellen und Kookkurrenzen finden, historische, soziale und regionale Homonyme und Antonyme identifizieren sowie Parallelstellen und Hapax legomena (Begriffe, die in einem Text oder Korpus nur einmal vorkommen) ausmachen. Bei alledem handelt es sich zwar um

grundlegende Operationen, die in den Geisteswissenschaften seit Jahrhunderten durchgeführt werden, in dieser bislang nicht gekannten Breite und Tiefe werden sie jedoch erst durch digitale Wörterbücher und Bibliographien möglich. Die digitale Welt hat auch die der Gelehrsamkeit erweitert und vertieft.

Bald kamen digitale Editionen von Werken und historischen Korpora dazu, vom Œuvre Gustave Flauberts über die Gedichte Walt Whitmans, die Literatur von Frauen und die Drucke des 16. Jahrhunderts bis hin zum Familienregister Irlands oder der in keinem herkömmlichen Druck abbildbaren mittelalterlichen Überlieferung der Parzival-Handschriften. Solche Projekte haben das Geschehen in den Digital Humanities in den letzten drei Jahrzehnten in einem Ausmaß bestimmt, dass nicht wenige diese Forschungsrichtung mit Retrodigitalisierung und digitaler Editorik gleichsetzen. Die Perseus Digital Library des Altphilologen Gregory Crane stellt die wohl größte Sammlung griechischer und römischer Quellen für die Forschung zur Verfügung. Und die erste, 1999 gegründete deutsche Digital-Humanities-Fachzeitschrift, herausgegeben von Karl Eibl und Fotis Jannidis, erscheint unter dem Titel *Jahrbuch für Computerphilologie*. Kein Zufall, dass Google mit der Massendigitalisierung von Büchern begonnen hat. Mehr Philologie war nie. Das hat die Digital Humanities nachhaltig geprägt.

Die digitalen Editionen sind unverändert das Brot-und-Butter-Geschäft, in den letzten Jahren kamen dann allerdings immer mehr kulturelle Objekte hinzu. So führt die digitale Perseus-Bibliothek heute Tausende archäologische Funde wie Münzen, Vasen oder Statuen auf. Auch Bilder und musikalische Werke werden erfasst. Prominente Beispiele sind die Digitale Mozart-Ausgabe, die in Kooperation mit Hewlett-Packard entsteht, und das Projekt »Universal Leonardo« des Kunsthistorikers Martin Kemp, das das wissenschaftliche und künstlerische Werk Leonardo da Vincis erschließt. Die digitale Ausgabe zeigt

nicht nur einzelne Gemälde in einer Auflösung, die kein Museum bieten kann, sondern ermöglicht auch die Betrachtung tieferer Schichten, die durch Röntgen- und Infrarotaufnahmen zugänglich gemacht wurden (Kohle 2013). Die Vorteile, die solche Editionen für die Forschung haben, liegen auf der Hand. Man kann diese Entwicklung mit Fug und Recht als Revolution betrachten, nicht zuletzt, weil Werke aus den Elfenbeintürmen der Wissenschaft geholt werden. Die Digitale Mozart-Ausgabe wird weltweit täglich von Hunderttausenden genutzt, die Grundlagen von Kennerschaft und Reputation ändern sich, weil nun jeder Oberstufenschüler auf Renaissance-Manuskripte zugreifen kann und jede Bachelor-Studentin Vorarbeiten zu schwer zugänglichen Bildern für ihre Studien heranziehen kann. So gerät auch unser Begriff von Kultur in Bewegung. Das Internet der kulturellen Dinge erwacht (Uckelmann et al. 2011).

II.

Eine wichtige Voraussetzung für den Erfolg der Digital Humanities und ihrer digitalen Editionen war die Standardisierung der Daten und Metadaten. Im Rahmen der Text Encoding Initiative (TEI) wurden über Jahrzehnte hinweg Vokabulare für digitale Auszeichnungsstandards entwickelt, die es ermöglichen, Texte unabhängig von der jeweils benutzten Soft- und Hardware eindeutig zu annotieren, also zu sagen, dies ist eine Gedichtzeile, das ist eine Sprecherfigur im Drama, dies ist der Name des Autors, das ist das Jahr der Erstveröffentlichung und vieles andere mehr. So entstand ein flexibles Vokabular, das für unterschiedliche Textsorten und Dokumenttypen in allen Sprachen geeignet ist. Erst solche in den Kopf der Dokumente und in die Dokumente selbst eingefügten Auszeichnungs-Tags machen aus einer Folge von Zeichen einen strukturierten Text.

Andere Disziplinen folgen, die Musikwissenschaft etwa mit der Music Encoding Initiative (MEI). Vor allem Bibliotheken, Rechenzentren und auch das World Wide Web Consortium sind auf solche Standards insbesondere für Metadaten angewiesen, um die digitalen Ausgaben in ihren Katalogen, auf Servern und Websites vorhalten und vernetzen zu können. Daten-Standardisierung und die Zusammenarbeit vieler Institutionen in aller Welt charakterisieren daher die digitalen Geisteswissenschaften. Digital Humanities, so könnte man sagen, sind international – oder sie sind nicht.

Trotz all dieser Erfolge wurde in den letzten Jahren wiederholt moniert, die Digital Humanities blieben hinter ihren Möglichkeiten zurück, seien zu sehr am Einzeltext sowie an der digitalen Edition orientiert und kultivierten in den historisch-hermeneutischen Traditionen ein enges Selbstverständnis (Thaller 2012), das man mit der eingängigen Formel des amerikanischen Dichters, Literatur- und Agrartheoretikers John Crowe Ransom als das des »close reading« bezeichnen kann. Gemeint ist damit die detaillierte Analyse der syntaktischen Verbindungen und semantischen Vieldeutigkeiten der großen Werke, und zwar Absatz für Absatz, Satz für Satz und Wort für Wort. Das verpflichte die Geisteswissenschaften jedoch zu sehr auf Kanon und Werk, auf das Seltene und künstlerische Modelle der Weltauslegung. Kontexte, die Pragmatik und überhaupt die Kulturgeschichte der Bedeutungen gerieten somit in den Hintergrund.

Wohl niemand hat diese traditionelle Herangehensweise so pointiert kritisiert wie der italienisch-amerikanische Komparatist Franco Moretti. In seinem vieldiskutierten Aufsatz »Conjectures on world literature«, der im Jahr 2000 in der *New Left Review* erschien, geht es ihm weniger darum, konzeptionelle Probleme der Digital Humanities zu lösen, er sucht vielmehr nach neuen Lösungen für sein Fach. Wie kann man eine Geschichte der Weltliteratur schreiben, wenn niemand auch nur

annäherungsweise alles lesen kann, schon gar nicht in allen Sprachen? Als Antwort präsentiert er sein Konzept des »distant reading«. »Literaturgeschichte«, schreibt Moretti dazu, »wird bald schon etwas anderes sein als das, was sie jetzt ist. Sie wird zu einer Literaturgeschichte aus ›zweiter Hand‹ werden, zu einem Patchwork aus den Ergebnissen anderer Forscher, und zwar *ohne die direkte Lektüre einzelner literarischer Texte*. Das ist immer noch ein ambitioniertes Unternehmen, im Prinzip ist es sogar ambitionierter als zuvor (Weltliteratur!); der Ehrgeiz ist jetzt *direkt proportional zum Abstand vom Text*. Je ambitionierter das Projekt, desto größer muss der Abstand sein.« (Moretti 2000)

Damit bricht Moretti so ziemlich mit allem, was die Literaturwissenschaft (und nicht nur sie) klassischerweise ausmacht, und tatsächlich haben Kollegen wie Stanley Fish nicht an Kritik am ehrgeizigen Projekt des distanzierten Lesens gespart (vgl. Erlin/Tatlock 2013). Dennoch wurde die Formel bald aufgegriffen. Sie fasst schließlich sehr gut zusammen, vor welcher Herausforderung wir angesichts der digitalen Revolution stehen: der Millionen-Bücher-Situation. Digital Humanities waren notgedrungen auf das einzelne Werk und den Kanon fokussiert, solange größere historische Sammlungen nicht ausreichend gespeichert und kontextualisiert werden konnten. Noch Ende des 20. Jahrhunderts markierte das *Grimmsche Wörterbuch* auf einer CD die Grenze des Möglichen. Mit dem Internet, der Verbilligung des Speicherplatzes und der rasant wachsenden Zahl der Digitalisate ist Big Data nun auch in den Geisteswissenschaften mehr als nur eine Metapher. Schätzungen von Google zufolge wurden weltweit bislang circa 130 Millionen Bücher gedruckt, fast 20 Millionen davon hat das Unternehmen mittlerweile digitalisiert. Als Google Books 2004 startete, hätte sich wohl niemand träumen lassen, dass nicht einmal zehn Jahre später so viele Bücher mit nur einem Mausklick erreichbar sein würden. In anderen Sammlungen wie beispielsweise Early English Books

online, Evans oder Gallica, im Deutschen Textarchiv oder im TextGrid-Repository liegen inzwischen ebenfalls Millionen Bilder, Drucke und Bücher vor. Wer die zweite Auflage eines anthropologischen Traktats aus dem 18. Jahrhundert sucht, ein persisches Manuskript, eine Notenhandschrift oder ein Porträt Michel de Montaignes aus dem 16. Jahrhundert, wird in den digitalen Archiven in Sekundenschnelle fündig. Doch auch ganz andere Ereignisse werden digital aufbereitet. So tragen Wissenschaftler, Politiker und Bürger im »Digital Archive of Japan's 2011 Disasters« alle möglichen medialen Spuren zusammen, die die Tsunami-Katastrophe hinterlassen hat: Bilder, Tweets, Nachrichten und politische Verlautbarungen. So entsteht ein Archiv der laufenden Weltereignisse.

Die digitalen Geisteswissenschaften vermessen die Kultur im buchstäblichen Sinn neu. Auch die Archäologie liefert dafür anschauliche Beispiele. Fanden Wissenschaftler einzelne Teile einer Skulptur, fertigten sie bislang Gipsabgüsse an, mit denen sie durch die Museen der Welt reisten, um zu prüfen, an welche Statue dieser Kopf oder jener Arm passt. Heute lassen sich diese Arbeiten per 3-D-Scanner erledigen. Was zuvor über unterschiedliche Kontinente verstreut war, wird nun mithilfe des Internets wieder zusammengefügt. Laserscanner erlauben es, Gelände (auch durch das Blattwerk der Bäume hindurch) mit einer nie gekannten Genauigkeit zu kartographieren. So entdeckten jüngst Damian Evans und Jean-Baptiste Chevance etwa 40 Kilometer nördlich der kambodschanischen Tempelanlage Angkor Wat die jahrtausendealte Stadt Mahendraparvata. Wo das bloße Auge nur Dschungel sieht, zeigt der Computer Details wie Verdichtungen im Boden oder Wasseransammlungen.

Historiker greifen ebenfalls auf die neuen digitalen Methoden zurück. So zielt beispielsweise die European Holocaust Research Infrastructure (EHRI) darauf ab, aus polnischen Steuer-

verzeichnissen, deutschen Deportationslisten und Zeugnissen wie Audio- und Filmaufzeichnungen die Namen der sechs Millionen Juden zu rekonstruieren, die während des Zweiten Weltkriegs ermordet wurden. Bislang ist lediglich circa ein Drittel der Namen bekannt. Die unterschiedlichen Quellen werden bearbeitet und dann in einer durchsuchbaren Datenbank zusammengefügt. Die schiere Datenmenge solcher und ähnlicher Vorhaben braucht den Computer. Solche Digital-Humanities-Projekte sind mehr als nur eine Fortsetzung der Geisteswissenschaften mit den Mitteln der Informationstechnologie. Sie eröffnen vielmehr neue Forschungswelten.

III.

Die aktuelle Transformation der Geisteswissenschaften betrifft weniger die Tiefe, in der Kultur mithilfe neuer Ansätze erschlossen werden kann, sondern die Breite der entsprechenden Unterfangen. Im analogen Zeitalter waren kulturellen Beständen enge Grenzen gesetzt, weshalb man sich auf einen abgeschlossenen Kanon konzentrieren musste. Das digitale Zeitalter macht nun (fast) die gesamte Kultur zugänglich. Kulturelle Evolution statt Geschichte – so könnte man zugespitzt die Richtung umschreiben, welche die geisteswissenschaftliche Forschung in den nächsten Jahren bestimmen wird. Plötzlich können wir nachvollziehen, wie sich eine bestimmte Keramiktechnik über große Kultur- und Zeiträume hinweg verbreitet hat. Wir können herausfinden, ob deren Ausbreitung jener ansteckender Krankheiten ähnelt, ob sie entlang von Handelsrouten erfolgte usw. Der niederländische Kulturhistoriker Joep Leersen wollte wissen, über welche Netzwerke die Idee des Nationalstaats im 19. Jahrhundert diffundierte und wie viele Korrespondenzpartner daran beteiligt waren. Die Ergebnisse visualisierte er in einem

Diagramm, das Edinburgh ebenso umfasste wie Sarajevo. So konnte er zeigen, dass nur wenige Autoren und Knotenpunkte notwendig werden, um der Idee Geltung zu verschaffen. So unterschiedliche Digital Humanists wie John Burrows, Julia Flanders, David Hoover, Fotis Jannidis oder Matthew Jockers (Jockers 2013) gehen unter anderem diesen Fragen nach: Sind Epochen nur historiographische Konstruktionen? Wann wird die Literatur von irischen Einwanderern zu einer amerikanischen? Welche neuen Themen treten um 1800 in Erscheinung? Schreiben Frauen anders als Männer? Schreibt Kleist ganz anders als die klassischen und romantischen Autoren seiner Zeit? Die Antworten basieren auf großen Datenmengen, werden mithilfe von Statistiken errechnet und graphisch visualisiert. Insgesamt haben wir es dabei mit etwas zu tun, das man als historisches »Reality-Mining« bezeichnen könnte. All das erweitert die Möglichkeiten der Geisteswissenschaften ungemein.

Und doch setzen die Geisteswissenschaften auch als Digital Humanities historische und hermeneutische Kenntnisse voraus und sie bedürfen weiterhin der theoretischen Modellierung. Denn was überhaupt als Idee oder Kulturtechnik gelten soll, ergibt sich nicht allein aus den Daten. Theorien der Evolution sind so notwendig wie solche der Geschichte. Daten sind eben noch keine Informationen, und Informationen brauchen eine Struktur, die sich wiederum nicht aus den Daten ableiten lässt. Deshalb geht es in den Digital Humanities derzeit gerade um Theorien, die die Vielzahl kultureller Artefakte und Zeugnisse in semantischen Ontologien ordnen können. Das Europäische Datenmodell der Europeana-Bibliothek oder das Resource Description Framework (RDF) basieren zum Beispiel auf semantischen Webstandards, in deren Entwicklung eine Unmenge kulturellen und theoretischen Wissens eingegangen ist. Will man die digitalisierten Dinge zueinander in Beziehung setzen, muss man sie mit Kontextinformationen versehen: Dieses Do-

kument stammt aus diesem kulturellen Raum und steht in einer Verbindung mit diesen Leserschaften; diese Karikatur hat etwas mit dem Boxeraufstand in China zu tun; Aphra Behn war die erste englische Berufsschriftstellerin; es gibt einen Zusammenhang zwischen dem Kolonialismus und den aktuellen Konflikten im Nahen Osten usw. usf. – kein Ende der Theorie und des Wissens, nirgends. Gerade weil datengetriebene Wissenschaften immer wichtiger werden, gerade weil Beobachter wie Jim Grey und Tony Hey von der Microsoft-Forschungsabteilung bereits von einem vierten Paradigma der datenintensiven wissenschaftlichen Erkenntnis sprechen, das auf die Paradigmen der Empirie, der Theorie und der Simulation folge (Hey et al. 2010), werden methodische Kritik und theoretische Modellierung wichtiger. Je mehr Daten, desto mehr Theorie, Wissen und Methodenkritik sind notwendig.

IV.

Für die digitale Modernisierung der Geisteswissenschaften ist die Kritik der Methoden und der Theorien immer auch eine algorithmische Kritik, eine Kritik der in Anschlag gebrachten rechnenden Verfahren (Ramsay 2011). Die neuen Verfahren werden schließlich nicht von Rechenmaschinen entwickelt, und sie ergeben sich auch nicht aus dem Material selbst. Womit kann man den Abstand zwischen Texten besser messen, mit Burrows Delta oder dem Euklidischen? Welche phylogenetische Methode aus der Biologie ist am besten geeignet, um stammesgeschichtliche, langfristige kulturelle Trends aufzuzeigen? Welche Vorteile bietet das SplitsTree-Verfahren gegenüber Dendrogrammen, wenn kulturelle Verwandtschaften dargestellt werden sollen? All diese methodischen und theoretischen Fragen halten erst mit den Digital Humanities Einzug

in die Geisteswissenschaften. Und sie brauchen Theorie. Die Auseinandersetzung mit den Algorithmen ist dabei allerdings nur Teil einer allgemeineren kritischen Diskussion über Fragen der digitalen Modernisierung. Und das hat mit dem Umstand zu tun, dass Digital Humanities nur möglich sind, wenn die Wissenschaftler auf die entsprechenden Daten zugreifen können (vgl. dazu die Beiträge von danah boyd und Kate Crawford sowie David Weinberger in diesem Band). Wer die Tagebücher Emma Darwins mit einem Stachelschweinstachel aus der Sammlung Johann Friedrich Blumenbachs in Beziehung setzen will, um die Vorgeschichte des Darwinismus zu erforschen, wer die Bautechnik frühneuzeitlicher Synagogen untersuchen oder verfolgen will, wie sich die Farbgebung auf Gemälden über die europäische Kunstgeschichte hinweg verändert hat, muss zuallererst einmal Zugang zu den einschlägigen Datenbanken haben (Zhou 2012). Wir benötigen also eine Kultur des Teilens, stattdessen dominieren derzeit jedoch Kommerzialisierung und Oligopole. Google Books ist hier nur ein Beispiel. Zwar verfügt das Unternehmen über die größte Menge digitalisierter und maschinell lesbarer Bücher, rund um diese Bibliothek gibt es allerdings diverse Geheimverträge. Museen verlangen Gebühren, wenn man ihre Stücke für die wissenschaftliche Arbeit nutzen möchte. Verlage fordern Phantasiepreise, wenn man Aufsätze miteinander in Beziehung setzen will, und verbieten die dauerhafte Einstellung von E-Books in öffentliche Bibliotheken. Staaten streben nach der digitalen Deutungshoheit über ihr nationales Erbe. Auch Wissenschaftler verhalten sich an dieser Stelle nicht immer vorbildlich (Lauer 2011). Während ganze Kontinente neuen Wissens auftauchen und das Wissen und erst recht die Wissenschaft neue Formen annimmt (Floridi 1995), verstärkt sich die Tendenz, Daten unter Verschluss zu halten. Große, kommerzielle Wissensaggregatoren wie Web of Science haben öffentliche Projekte wie das Internet Archive oder Euro-

peana längst überholt. Schlagworte wie »Crowd Science« oder »Citizen Science« (Blanke 2014) bleiben einstweilen nur schöne Versprechungen, weil die Daten-Kleinstaaterei die Kultur mit Zollschränken belegt. Kultur- und Wissenschaftspolitik werden daher in Zukunft immer auch Politik des Digitalen sein, und die Digital Humanities werden sich mit Initiativen wie Open Access und Creative Commons verbünden müssen (Lessig 2006). Theoretisch mag ein mittelalterlicher japanischer Roman nur einen Tastendruck entfernt sein, Roboter wären jederzeit in der Lage, diesen Text in fast alle Sprachen zu übersetzen, Algorithmen könnten kulturelle Zusammenhänge aufschlüsseln usw. – doch die Daten sind zu selten frei zugänglich. Die sich abzeichnende Feudalstruktur der digitalen Gesellschaft und die Mentalität der Digital Humanities sind insofern nicht kompatibel. Dennoch: Die digitale Vermessung der Kultur hat soeben begonnen, und wir können sagen, wir sind dabei.

Literatur

- Blanke, Tobias (2014), *The Ecosystem of Digital Assets. Crowds and Clouds*, Oxford: Chandos.
- Daston, Lorraine/Katharine Park (2002), *Wunder, Beweise und Tatsachen. Zur Geschichte der Rationalität*, Frankfurt am Main: Fischer.
- Erlin, Matt/Lynne Tatlock (Hg.) (2013), *Distant Readings. Descriptive Turns. Topologies of German Culture in the Long Nineteenth Century*, Rochester: Camden.
- Floridi, Luciano (1995), »Internet. Which future for organized knowledge, Frankenstein or Pygmalion?«, in: *International Journal of Human-Computer Studies* 43/2, S. 261-274, online verfügbar unter: {<http://www.philosophyofinformation.net/publications/pdf/frank.pdf>} (Stand August 2013).
- Grzybek, Peter (Hg.) (2007), *Contributions to the Science of Text and Language. Word Length Studies and Related Issues*, Dordrecht: Springer.
- Hey, Tony/Stewart Tansley/Kristin Tolle (Hg.) (2010), *The Fourth Para-*

- digm. Data-Intensive Scientific Discovery*, Seattle: Microsoft Research, online verfügbar unter: {<http://research.microsoft.com/en-us/collaboration/fourthparadigm/>} (Stand August 2013).
- Jockers, Matthew L. (2013), *Macroanalysis. Digital Methods and Literary History*, Urbana/Chicago/Springfield: University of Illinois Press; zusätzliches Material ist online verfügbar unter: {<http://www.matthewjockers.net/macroanalysisbook/>} (Stand August 2013).
- Joula, Patrick (2008), *Authorship Attribution*, Boston/Delft: Now.
- Kirschenbaum, Matthew (2010), »What is digital humanities and what's it doing in English Departments?«, in: *ADE Bulletin* 150, S. 1-7, online verfügbar unter: {http://mkirschenbaum.files.wordpress.com/2011/01/kirschenbaum_ade150.pdf} (Stand August 2013).
- Kohle, Hubertus (2013), *Digitale Bildwissenschaft*, Glückstadt: Hülsbusch, online verfügbar unter: {<http://archiv.ub.uni-heidelberg.de/artdok/2185/>} (Stand August 2013).
- Lauer, Gerhard (2011), »Bibliothek aus Daten«, in: *Kodex* 1, S. 79-85, online verfügbar unter: {http://gerhardlauer.de/index.php/download_file/view/187/85/} (Stand August 2013).
- Lessig, Lawrence (2006), *Freie Kultur. Wesen und Zukunft der Kreativität*, München: Open Source Press, das englische Original ist online verfügbar unter: {http://www.jus.uio.no/sisu/free_culture.lawrence_lessig/portrait.letter.pdf} (Stand August 2013).
- Lunenburg, Peter/Anne Burdick/Johanna Drucker/Todd Presner/Jeffrey Schnapp (2012), *Digital Humanities*, Cambridge (Massachusetts): MIT Press, online verfügbar unter: {http://mitpress.mit.edu/sites/default/files/titles/content/9780262018470_Open_Access_Edition.pdf} (Stand August 2013).
- Moretti, Franco (2000), »Conjectures on world literature«, in: *New Left Review* 1, S. 54-68, online verfügbar unter: {<http://newleftreview.org/II/1/franco-moretti-conjectures-on-world-literature>} (Stand August 2013).
- Ramsay, Stephen (2011), *Reading Machines. Towards an Algorithmic Criticism*, Urbana/Chicago/Springfield: University of Illinois Press.
- Thaller, Manfred (Hg.) (2012), *Controversies Around the Digital Humanities*, Sonderheft der Zeitschrift *Historical Social Research* (37/3).
- Uckelmann, Dieter/Mark Harrison/Florian Michahelles (2011), *Architecting the Internet of Things*, Heidelberg/New York/Dordrecht/London: Springer.
- Unsworth, John (2000), »Scholarly primitives: what methods do hu-

manities researchers have in common, and how might our tools reflect this?«, online verfügbar unter: {<http://people.lis.illinois.edu/~unsworth/Kings.5-00/primitives.html>} (Stand August 2013).

Zhou, Mingquan/Guohua Geng/Zhongke Wu (2012), *Digital Preservation Technology for Cultural Heritage*, Heidelberg/New York/Dordrecht/London: Springer.