

## 1 Geschichte der Digital Humanities

### 1.1 | Texte und Informationstechnologie: Der Gründungsmythos der Digital Humanities

Informationstechnologien werden heute in allen Fächern der geisteswissenschaftlichen Fakultäten und Fachbereiche verwendet. Die frühesten Beispiele dafür finden sich in den Fächern, die sich mit Texten beschäftigen, sei es im sprach- oder literaturwissenschaftlichen Sinn. Aus den Anfängen der Verwendung der frühen Computertechnologie im Umgang mit Texten sticht vor allem eine Geschichte hervor, die geradezu als »Gründungsmythos« der Digital Humanities gelten kann. 1940 arbeitete ein junger Jesuit, **Roberto Busa**, an einer Dissertation zum Thema des Präsenzbegriffes bei Thomas von Aquin. Sehr bald musste er feststellen, dass das offensichtliche, die Suche nach dem Wort »presentia«, wenig brachte, da das zu Grunde liegende intellektuelle Konzept meist nur indirekt durch die Verwendung der Präposition »in« angezeigt wurde. Als Funktionswort tritt »in« nicht selten auf und fehlt in allen Registern der gedruckten Edition der Texte; die einzige Möglichkeit bestand also darin, Tausende von Druckseiten Zeile für Zeile und Wort für Wort durchzugehen und, sobald »in« gefunden war, festzustellen, ob seine Verwendung in diesem Fall auf das untersuchte Phänomen verweise.

10.000 handgeschriebene Karteikarten und einen Weltkrieg später verteidigte Padre Busa 1946 seine Dissertation erfolgreich – und war von seinen Ordensoberen dazu bestimmt, seine Laufbahn weiterhin dem Denken des Aquinaten zu widmen. Zumeist davon überzeugt, dass seine Konzeption, dass die Interpretation der einem Texte zu Grunde liegenden Philosophie besser auf dem Studium unbewusst verwendeter Worte und Konstruktionen aufbauen solle, als auf der Analyse der offensichtlich bedeutungstragenden, schlug er noch 1946 vor, dafür eine **Konkordanz**, also ein Verzeichnis *aller* bei **Thomas von Aquin** vorkommenden Worte, jeweils begleitet von einem kurzen Stück Kontext, zu produzieren, um diesen Ansatz für die weitere Forschung systematisch zu unterstützen. Dass dies mit herkömmlichen Mitteln nur wenig erfolgversprechend war, hatten ihn die 10.000 Karteikarten gelehrt.

Auf der Suche nach geeigneten Technologien gelang es ihm 1949 schließlich, IBM, damals vor allem als Hersteller von Lochkartensortiermaschinen profitabel, für die technische und finanzielle Unterstützung des Vorhabens zu gewinnen. Er war nicht nur ein hervorragender Wissenschaftler, sondern auch ein gewiefter Forschungspolitiker, und konnte so bald danach schließlich darauf verweisen, dass Remington Rand – ein Konkurrent von IBM – einen anderen Forscher bereits mit seinen Computern und Magnetbändern bei der ähnlichen Aufgabe der Erstellung einer Bibelkonkordanz unterstützt habe. Was dazu führte, dass auch der Aquinat von diesem Zeitpunkt an von den Lochkartensortierern auf die ersten Generationen von IBM Computern transferiert wurde.

Abgesehen davon, dass die Geschichte eines (bis zu seinem Lebensende) von Charisma und Charisma übersprühenden jungen Forschers, der einen der einflussreichsten Industriellen seiner Zeit davon überzeugte, einem genuin geisteswissenschaftlichen Vorhaben, der Untersuchung der Gedankenwelt Thomas von Aquins, seine uneingeschränkte Unterstützung zu leihen, einen hervorragenden Einstieg in die Geschichte einer Disziplin darstellt, versteckt sich darin ein leicht übersehbarer Zwiespalt, der die Digital Humanities bis heute prägt. Offensichtlich wurde die frühe

Rechnertechnologie hier als ein **Werkzeug** verwendet, das den Zeitaufwand für die Bewältigung von Routineaufgaben signifikant verringerte. Padre Busa selbst betont in seiner Geschichte des ›Index Thomisticus‹, wie die Konkordanz genannt wurde, aber gar nicht so sehr den Nutzen des Werkzeugs: ›Ihn fasziniert in diesem Aufsatz viel mehr die Erkenntnis, dass die Konzentration auf unbewusste sprachliche Merkmale – Funktionswörter, also: ›in‹ – im Gegensatz zu den offensichtlich bedeutungstragenden, einen **methodisch** völlig neuen Zugang zu der in einem Text ausgedrückten Gedankenwelt ermöglicht, per Definition also einen Zugang, der über die einzelne Studie hinausgeht.

In der Tat hat die Beschäftigung mit den nicht-intentionalen sprachlichen Merkmalen in den Digital Humanities eine zweite Wurzel etwa zur selben Zeit. In der Geschichte der amerikanischen Verfassung sind eine Serie politischer Essays – die »**Federalist Papers**« –, die unter einem gemeinsamen Pseudonym von Alexander Hamilton, James Madison, und John Jay zur Zeit der Debatte über ihre Ratifizierung geschrieben wurden, von großer Bedeutung. Festzustellen, wer der tatsächliche Verfasser eines kleinen Teils davon war, bei dem nur das Pseudonym, nicht der wirkliche Autor bekannt war, galt lange Zeit als eine der spannenderen Fragen der amerikanischen Verfassungsgeschichte – nicht so sehr, weil daran die Interpretation der Verfassung gegangen hätte, wohl aber, weil sich hier an einem Detailproblem zeigen lässt, was die Kritik historischer Texte herausfinden kann und was nicht.

Schon 1940/1941, in etwa zu der Zeit, als Padre Busa seine Karteikarten füllte, beschäftigte sich der amerikanische Statistiker Frederick Mosteller, zusammen mit Frederick Williams mit der Frage, ob man das beschriebene Rätsel durch wahrscheinlichkeits-theoretische Ansätze klären könne. Ein Ansatz, der damals durchgebrochen, nach der Verfügbarkeit der ersten Rechnergenerationen jedoch von Frederick Mosteller, nunmehr zusammen mit David L. Wallace, zu einer Untersuchung ausgebaut wurde, die 1964 als *Inference and Disputed Authorship: Federalist Papers* erschien und gemeinhin als Ausgangspunkt der rechnergestützten Beschäftigung mit Fragen des Stils und der Verfasserschaft gilt, die sich ebenfalls auf **nicht bewusste Sprach-eigen-tlichkeiten**, Funktionswörter, stützt und womit sich die Verbindung zu Roberto Busa schließt. Beziehungsweise zur Erschaffung, dass arbeitstechnische Erleichterungen die Anwendung neuer methodischer Herangehensweisen ermöglichen.

Wir haben hier also Beispiele dafür, wie die neuen technologischen Vorgehensweisen erlaubten, die schon früher erwogen worden waren, aber als zu aufwändig nicht umgesetzt werden konnten. Wobei es Versuche, allgemeine Einschätzungen durch die penible Untersuchung für sich gesehen trivial wirkender Details zu untermauern auch in anderen Disziplinen schon früher gab – ein Beispiel dafür bilden die Versuche des Kunsthistorikers Giovanni Morelli im 19. Jahrhundert, allgemeine stilistische Aussagen durch den peniblen Vergleich des Umgangs mit häufig wiederkehrenden, etwa anatomischen, Details in Bildern zu untermauern. Dementsprechend wäre es auch falsch zu glauben, dass die Beschäftigung mit der neuen Rechnertechnologie in den Geisteswissenschaften sich ausschließlich auf die Untersuchung sprachlicher Phänomene bezogen hätte: In der wohl ersten internationalen von Dell Hymes organisierten Konferenz des entstehenden Forschungsfeldes 1962 auf der Burg Warrenstein, mit dem erst 1965 erscheinenden Konferenzband *The Use of Computers in Anthropology*, war bereits die ganze Bandbreite der geisteswissenschaftlichen Disziplinen vertreten, wobei sich die Anwendungen zwischen der Nutzung statistischer Herangehensweisen und dem analytischen Herangehen an Texte in etwa gleich verteilten.

## 1.2 | Eine Community entsteht: Die frühen Jahre

Der Zweck dieses Bandes ist nicht wirklich klar, welche Konsequenzen die Verfügbarkeit der Rechnertechnologie – aufbauend auf Busas Beispiel, aber auch häufig in Unkenntnis dessen –, in den 1950er Jahren in den Geisteswissenschaften insgesamt hatte. Knapp nach der Konferenz von 1962 wurden jedoch die Vortragsbände mehrerer einschlägiger Konferenzen sichtbar, die eine erstaunlich stark entwickelte Forschungsszene zeigen: Es gab zahlreiche Konkordanzprojekte und auch erste Zentren für die systematische Anwendung der elektronischen Technik, etwa in Cambridge, wo Roy Wisbey 1964 ein *Linguistic Computing Centre* gründete, das sich seinen eigenen Forschungsinteressen entsprechend überwiegend dem Mittelhochdeutschen widmete. In der Archäologie gab es große, auf sehr intensive statistische Analysen ausgerichtete Fachkongresse. Sobald das Bild klarer wird, zeigte es bereits eine erstaunliche Vielfalt, eine Community von Forschern, die einander auf den Konferenzen dieser ersten Welle aber wohl überwiegend überrascht darüber begegneten, wie nach die Szene schon war.

Diese Überraschung ist nicht ganz unbegründet: Nicht nur musste selbst die Unterscheidung zwischen Groß- und Kleinbuchstaben auf Lochkarten mit ziemlichem Aufwand dargestellt werden; ein wesentliches Problem wird klar, wenn wir die Einleitung des Herausgebers der Tagung von 1962 lesen: Nachdem moderne **Macrosamplers** jetzt den Zugriff auf den Maschinencode unnötig gemacht hätten, sei ein Benutzerkomfort bei der Programmierung erreicht, der die zusätzlichen Vereinfachungen durch Höhere Programiersprachen nicht mehr unbedingt erforderlich mache. Was damals nicht ganz so obskur erschien wie heute, das Problem fehlender **weiterer** Software jedoch sehr gut indirekt beschreibt.

Die weitere Entwicklung war durch zwei wichtige Schritte gekennzeichnet: Einerseits organisierte sich die wissenschaftliche Community der Geisteswissenschaftler, die die Informationstechnologie nutzten, selbst, so dass der Informationsaustausch vereinfacht wurde. Seit 1966 erschien, lange Jahre als zentrales Informationsmedium des Faches, unter der Herausgeberschaft von Joseph Raben die Zeitschrift *Computers and the Humanities* (CHUM), deren Bedeutung, später, als sie in kommerzielle Hände gefallen war, und seit 2005 als *Language Resources and Evaluation* erschien, freilich stark abgenommen hat. Damals war sie jedoch ein zentrales Medium, das andererseits maßgeblich mitgeholfen hat, **Fachgesellschaften** ins Leben zu rufen (In Europa seit 1973 die »Association for Literary and Linguistic Computing«, etwas später, 1978, die amerikanische »Association for Computers and the Humanities«).

Durch diese Verbände entstanden auf der Basis jährlicher Konferenzen Foren, in denen die einzelnen – wie wir ja gesehen haben, sehr zahlreichen – Ansätze verglichen werden konnten, was die Möglichkeit zur Entwicklung gemeinsamer Ansätze, über die Nützlichkeitserwägungen eines einzelnen Projekts hinaus bot. Bei ähnlichen **Konferenzserien**, wie der ebenfalls 1973 ins Leben gerufenen und ungenießerisch »Computer Applications and Quantitative Methods in Archaeology« (CAA), wird der tragende Verein hinter der Konferenzserie als solcher gar nicht sichtbar. Und diese Verbände fungieren auch als Träger von **Zeitschriften**, die die Gemeinsamkeiten des Umgangs mit Informationstechnischen Methoden, jenseits der Spezifika des jeweiligen disziplinären Kontexts, schärfer zeigen, als dies in den Konferenzen anderer Publikationen möglich wäre: So die Zeitschrift *Literary and Linguistic*

Computing seit 1973 (seit 2015 fortgeführt als *Digital Scholarship in the Humanities*) oder die reine Internetschrift *Digital Humanities Quarterly*, seit 2007 unter <http://www.digitalhumanities.org/dhq/> erreichbar und unter der Ägide der Alliance of Digital Humanities Organization (ADHO) erscheinend, die sich bemüht, die Tätigkeit der älteren Verbände zu bündeln.

### 1.3 | Die Welt wird einfacher – Programmpakete

Dieser kurze Vorgriff auf die Gegenwart darf freilich keinesfalls den Blick auf die andere, noch entscheidendere Entwicklung im Umgang von Geisteswissenschaftler mit der Informationstechnik in den späten 1960er und frühen 70er Jahren verstellen: Die Verfügbarkeit von **Programmpaketen**, die den Zugang auf bestimmte technische Verfahren, als Implementation bestimmter methodischer Ansätze, entscheidend vereinfachte. Aus den Bemühungen von Wilhelm Ott in Tübingen, zunächst Assemblerfunktionen bereitzustellen, die den Umgang mit Texten in der Programmierung erleichtern sollten, entstand »TU-STEP«, das für die Umsetzung gedruckter Editionen auf dem Rechner den entscheidenden Schritt von der Formulierung von Auswertungswünschen in einer auf das technologische Werkzeug ausgerichteten höheren Programmiersprache in eine an den Auswertungswünschen orientierte Scriptsprache bedeutete. Wobei die Tatsache, dass neben der Vorbereitung der Edition für den Druck auch zahlreiche vorbereitende technische Dienste und Analysemöglichkeiten zugänglich wurden, von höchster Bedeutung war.

In anderen Fällen war der Unterschied in der Komplexität der Programmierung in einer Programmiersprache und der in der Kommandosprache eines Programmpakets noch drastischer: Anspruchsvolle statistische Analysen durchzuführen, setzte zunächst erhebliche mathematische und Programmierkenntnisse voraus. Mit der Verfügbarkeit statistischer Anwendungssoftware, wie dem »Statistical Package for the Social Sciences« (seit 1968), wurde es möglich, komplexe statistische Berechnungen durchzuführen, ohne die Verfahren zu verstehen, was die Anzahl **quantitativer Untersuchungen** in den frühen 1970er Jahren, quer durch alle Disziplinen der Geisteswissenschaften, buchstäblich explodieren ließ. Das Wort Emmanuelle Roy Laduries, eines der prominentesten französischen Historiker des vorigen Jahrhunderts »dans ce domaine au moins, l'histoire de demain sera programmée ou il ne sera plus« (*Le Nouvel Observateur*, 8 mai 1968), greift hier nur voraus. Freilich: Das Geschick der Anwendbarkeit von Verfahren, die man nicht verstehen musste, war auch gefährlich und manche der frühen quantitativen Analysen enthalten untreueiche Kunstfehler.

Diese zu vermeiden, war bei den Programmpaketen einfacher, die direkt aus der entstehenden Community, welche sich aus Geisteswissenschaftlern zusammensetzte, die Rechner verwendeten, abstimmen – das »Oxford Concordance Program« (OCP), die »Text Analysis Computing Tools« (TACT) oder »WordCruncher« versuchten, die aus der Buss'schen Tradition entstandenen Konkordanzansätze ohne die Notwendigkeit eigener Programmierarbeiten zu unterstützen, freilich meist in gewisser Hinsicht trivialisiert. Während ein wesentlicher Bestandteil der Thematikkonkordanz ein aufwändiges Verfahren war, die Wortformen – *facto, fecit, factum, ...* – auf Wortstämme oder Lemmata – *facere* oder *fac* – zurückzuführen, damit man nicht von nicht bedachten Flexionsproblemen in die Irre geführt werde, hat sich die große Mehrzahl im weitesten Sinne konkordanzorientierter Projekte hier meist wesentlich bescheidener gegeben. In der Tat kam es in den 1970er und 80er Jahren auch zu et-

ner deutlichen Trennung zwischen dem Computereinsatz in den Sprachwissenschaften und allen anderen Geisteswissenschaften, einschließlich der literaturwissenschaftlichen Disziplinen: In der Sprachwissenschaft etablierte sich mit der **Computeringuistik** eine gut definierte akademische Disziplin, für die der Umgang mit morphologischen und grammatikalischen Fragen zentral war, während die Abgrenzung zwischen Computeranwenden und nicht anwenden unter den Wissenschaftlern in allen anderen Disziplinen sehr viel fließender war.

In der Tat sollte nie übersehen werden, dass die große Mehrzahl der Geisteswissenschaftler, die die neuen technischen Möglichkeiten einsetzten, dies **rein pragmatisch** getan haben: Sie haben sie als Werkzeuge verwendet, die den Zeitaufwand für die Bewältigung von Routineaufgaben signifikant verringern. Dementsprechend war auch der Boom der **quantitativen Verfahren** in den 1970er Jahren nicht so sehr vom Wunsch geprägt, methodisch innovativ zu sein, als vielmehr davon, sich in einer schwer überschaubaren Zahl von Quellenbelegen zu orientieren. Als dies durch die leichtere Verfügbarkeit von Datenbanken in den 1980er Jahren *ohne* statistische Kodierungen möglich war, da nunmehr relativ leicht zugängliche **Datenbanksysteme** zur Verfügung standen, ebnete die Begeisterung für quantitative Anwendungen rasch ab. Wie gewichtig dieser neue Trend war, lässt sich am deutlichsten dadurch belegen, dass Joseph Raben, den wir schon als Pionier kennengelernt haben, sich in den 1980er Jahren aus den Vorläufern der Digital Humanities weitgehend zurück zog und einige Jahre als Organisator einer eigenen Konferenzreihe »Databases in the Humanities and the Social Sciences« wirkte, die damals deutlich mehr Teilnehmer an zog, als die etablierten traditionellen Konferenzserien.

### 1.4 | Die Welt wird noch einfacher – der Personal Computer

Vereinfacht kann man sagen, dass sich damals die Vorläufer der heutigen Digital Humanities als jene Gruppe bildeten, die darauf bestanden, dass der Einsatz der jetzt leichter zugänglichen Technologien nicht nur auf den eines reinen Werkzeuges hinauslaufen sollte, sondern auch **methodische Weiterungen** und Überlegungen nach sich ziehen könne. Dieses Phänomen verschärfte sich durch den nächsten technologischen Durchbruch: Während bis Mitte der 1980er Jahre »Rechnereinsatz« den Zugang zu einem Großrechner erforderte, also oftmals logistisch gut geplante Ausflüge in ein oft universitäres Neubaugelände ausgelagertes Rechenzentrum, holte der Personal Computer die Rechenleistung plötzlich auf den Schreibtisch – zusammen mit **aus heutiger Sicht bescheidenen, damals aber völlig neuen graphischen Möglichkeiten**. Zunächst trennte das die »methodisch bewussten« arbeitenden Forscher eher von den zahlreichen gelegentlichen Anwendern. Die Verarbeitung von Texten – im Sinne der heutigen Textverarbeitung – war lange Zeit ein Spezialgebiet des Rechnereinsatzes in den Geisteswissenschaften, das sich mit den Verbindungen von aus wissenschaftlichen Ansprüchen her maschinenlesbar gemachten Texten zu **elektronischen Satzsystemen** beschäftigte. Als Spezialgebiet kollabierte dieser Bereich innerhalb kurzer Zeit: »Microsoft Word« sowie seine Vorläufer und verdrängten Konkurrenten verwandelten vergleichsweise arkanes Wissen über die Produktion von Texten in **eine allgemeine Kulturtechnik**.

Was nun selbstverständlich nicht heißen soll, dass die PC-Revolution der Verarbeitung der Rechner in den methodenbewussteren Teilen der Geisteswissenschaft-



ten geschadet hätte. Abgesehen davon, dass es für weite Bereiche der Geisteswissenschaften damit möglich wurde, technische Werkzeuge einzusetzen, ohne weiter darüber nachzudenken, abgesehen auch davon, dass natürlich die bisherigen Ansätze zur Erschließung von Texten oder zur quantitativen Analyse nochmals sehr viel einfacher einsetzbar wurden, als durch die Entstehung darauf spezialisierter Programmpakete, eröffneten sich völlig neue Möglichkeiten. Wie bereits erwähnt war die Vorbereitung komplexer Druckbilder lange Zeit ein Spezialgebiet, das hohen Stellenwert genoss. Dies gilt ganz besonders für den Bereich der **kritischen Editionen**, die anstreben, Texte in einer Form zugänglich zu machen, die schwierige Überlieferungssituationen in mehreren gereinigten Manuskripten (Textzeugen) verständlich macht und gleichzeitig einen authentischen Text bereitstellt. Meist mit mehreren Anmerkungsapparaten versehen, war der Druck dieser Werke immer extrem kostspielig und die erwähnte Tradition der elektronischen Erstellung des Drucksatzes versuchte weiterhin zu ermöglichen, was anders nicht mehr finanzierbar gewesen wäre.

Soweit ging es jedoch nur um eine Beibehaltung des Erreichten. Mit der Verbreitung der Personal Computer ging die Entwicklung des Konzepts digitaler oder **dynamischer Editionen** einher, die nicht mehr den Druck, sondern den Bildschirm des Rechners als Medium für die Darstellung des Textes ansehen.

Ansätze dazu gibt es seit der Frühzeit der PCs, und zwischen 1985 und den späten 1990er Jahren entstand mit den Arbeiten – neben vielen anderen – von Peter Robinson (vor allem: Canterbury Tales), Kevin Kiernan (Electronic Beowulf), Claus Huitfeldt (Wittgenstein Nachlass), Paolo D'Iorio (Hypermietzsche), Friedrich Aspetsberger (Musil Nachlass) und Walter Morgenstaler (Gottfried Keller) eine Tradition CD-basierter digitaler Editionen, die als einen inoffiziellen de facto-Standard etwa Folgendes erwartet: Ein Text kann wahlweise mindestens als (a) buchstabengetreue Transkription, mit Darstellung von Streichungen, Varianten etc. und (b) als geglätteter Lesetext so dargestellt werden, dass zu jeder Textstelle das dazugehörige Manuskript einsehbar ist bzw. die dazugehörigen Manuskripte einsehbar sind und der Text von einem ausführlichen Kommentar begleitet wird.

Natürlich gilt auch hier, dass nicht alle geisteswissenschaftlichen Projekte zwischen 1985 und der Jahrtausendwende digitale Editionen waren, genauso, wie nicht alle Projekte der 1970er Jahre mit quantitativen Verfahren und nicht alle Projekte der frühen 1980er Jahre mit Datenbanken gearbeitet haben. Sie sind aber besonders typisch für die Optionen, die sich durch die Nutzung der neuen technischen Möglichkeiten ergaben. Die digitalen Editionen beschreiben daher vielleicht am besten, was die wesentlichen Auswirkungen der PC-Technologie waren: Die Rechenleistung war unmittelbar am Arbeitsplatz verfügbar, und sie konnte auch für die Darstellung von Ergebnissen, nicht nur für die analytische Erzeugung dieser Ergebnisse, verwendet werden. In vielen Fällen führte dies dazu, dass wohlbekannte Produkte der rechnergestützten Geisteswissenschaften ihr Gesicht völlig wandelten: Die Konkordanz, mit der wir diese Geschichte von Begegnungen der Geisteswissenschaften mit der Informationstechnik begannen, existiert so nicht mehr. Freilich ist die abstrakte Aufgabe »untersuche alle Kontexte, in denen in einer Gruppe von Texten eine bestimmte Wortform vorkommt« alles andere als erledigt: Sie wird seit dem Ende des vorigen Jahrhunderts aber nicht mehr im Druck, sondern dynamisch erledigt. Sie wird nicht als vollständiges Verzeichnis aller möglichen Fundstellen im Druck eingefroren, sondern als Softwarepaket mit einem bestimmten Leistungsspektrum zusammen mit den dazu benötigten Daten ausgeliefert.

Darüber hinaus hat die Möglichkeit, den Rechner auch als **graphisches Präsentationsmedium** zu nutzen, natürlich ganze Anwendungsfamilien geschaffen: Die systematische Erfassung und Publikation kunsthistorischer Objekte in bildlicher Form auf CDs nahm an der Schwelle der 1980er zu den 1990er Jahren ihren Anfang. Die Vorstellung, PCs könnten als Lehrmittel eingesetzt werden, oft auf der Basis von **Lehrprogrammen**, die dieses Konzept der Kombination von softwaretechnischen Möglichkeiten mit einem darauf abgestellten, gerne auch multimedialen, Datenpaket für didaktische Zwecke einsetzen sollten, tat dies ebenso.

Um zu verstehen, wie sich daraus der nächste Schritt ergab, müssen wir zunächst noch einmal einen zurückgehen. Eine der größten Schwierigkeiten der frühen Stadien des Rechinereinsatzes in den Geisteswissenschaften war immer die Erzeugung der vom Rechner zu verarbeitenden Daten. In nicht ganz wenigen Fällen wiederholte sich in den frühen Jahren die Situation, dass nach mehrjähriger mit großer Hingabe durchgeführter **Dateneingabe** die Daten für die Analyse ungeteilt zu dem Zeitpunkt abgeschlossen vorlagen, zu dem auch die Finanzierung des Projekts endete. Dies besserte sich über die Zeit hinweg – an der Tastatur eines PCs einen Text abzutippen, braucht einen Bruchteil des Aufwandes, der benötigt wird, um ihn unter Kodierern aller Großbuchstaben, nationalen Sonderzeichen und sonstiger Besonderheiten (z. B. Längenbezeichnung der Vokale in mittelhochdeutschen Texten) auf Lochkarten zu übertragen.

So erfreulich dies ist, ändert es aber zunächst nichts an der Tatsache, dass manche Textmerkmale gekennzeichnet werden müssen: Stellt ein Wort einen Personennamen oder eine topographische Bezeichnung dar? So lange dies von jedem Wissenschafter abschließlich für die Zwecke der eigenen Studie gelöst wird, bleiben die Daten nur für diese Studie nutzbar. Wenn wir aber, nach dem oben Gesagten, davon ausgehen, dass Geisteswissenschaftler, die informationstechnische Möglichkeiten nutzen, nicht mehr nur die einmal gedruckten analytischen Ergebnisse als ihr Produkt verstehen, sondern die Verbindung aus programmtechnischen Möglichkeiten und dafür passendem Corpus, als wesentlich ansehen, liegt es ungemein nahe, danach zu suchen, wie man die Daten so vorbereiten kann, dass sie auch für andere Softwareanwendungen genutzt werden können. Und noch viel näher liegt die Idee, dass es nützlich sein könnte, eine bestimmte analytische Idee an bereits vorliegenden Daten auszuprobieren, bevor man selbst mehrere Jahre in die Dateneingabe investiert. Von der Möglichkeit, mit Hilfe der »Optical Character Recognition« (OCR) diese Dateneingabe massiv zu beschleunigen, bzw. den Umfang des analysierbaren Corpus massiv zu vergrößern, ganz abgesehen.

## 1.5 | Vernetzungen von Personen und Ressourcen

Ebenfalls Mitte der 1980er Jahre wandelte sich die Vorstellung von den Informations-technologien (IT) schieflich in die der »Informations- und Kommunikationstechnologien« (ICT) – anders ausgedrückt, mit Verbreitung der E-Mail etablierte sich der Rechner als **Kommunikationsgerät**, was für die Community der informationstechnisch arbeitenden Geisteswissenschaftler die Frage löste, wie man zwischen den Jahreskonferenzen den Kontakt aufrechterhalten könne, was zur Etablierung mehrerer Fachforen führte, von denen die Mailingliste »Humanistik«, ins Leben gerufen und betrieben von Willard McCarty seit 1987, am wichtigsten war und wohl auch immer noch ist.



Ungefähr zeitgleich mit der Verbreitung der Personal Computer ergab sich also der Wunsch, Daten soweit als möglich so aufzubereiten, dass sie wiederverwertbar und von den unmittelbar anstehenden analytischen Plänen, vor allem aber von den Anforderungen eines spezifischen zunächst ins Auge gefassten Softwaresystems, unabhängig werden sollten. Parallel dazu entstanden Möglichkeiten wesentlich enger kommunikativer Vernetzung mit den Kollegen. Daraus entwickelte sich das Konzept einer gemeinsamen Anstrengung der Community zur Verbesserung der Fähigkeiten, Daten auszutauschen und für möglichst viele Softwarelösungen verarbeitbar zu machen. Die **Text Encoding Initiative** entstand als Antwort auf diesen Wunsch im Spätherbst 1987 (s. Kap. 17.7 und 18.5). Für die weitere Entwicklung des damals generell als Humanities' Computing bezeichneten Feldes ist hauptsächlich wichtig, dass sie aus dem Verständnis entsprang, dass es projektübergreifende Herausforderungen von geisteswissenschaftlichen Projekten gibt, die die Informationstechnik nutzen und von den disziplinären Fragestellungen einzelner Vorhaben unabhängig sind, d. h. durch Anstrengungen unterstützt werden sollten, die ebenfalls die Einzeldisziplinen übergreifen.

Wir haben schon gesagt, dass die digitalen Editionen ein besonders gutes Beispiel für die neuen Möglichkeiten der Informationstechnologie mit Ausbreitung der PC-Technologie darstellen und dass es knapp vor der Jahrtausendwende ein Informelles, aber in sich konsistentes, Paradigma gab, wie man derartige Editionen gestalten sollte. Nichtsdestoweniger haben gerade Pioniere, wie Peter Robinson, Mitte der 2010er Jahre ein vergleichsweise deprimierendes Bild dieses Ansatzes gezeichnet. Hierfür ist ein der damaligen Vorgehensweise geschuldetes Problem verantwortlich: Gedruckte Editionen haben zahlreiche dem statischen Medium **inhärente Beschränkungen**. Kann man eine davon verwenden und hat verstanden, was ein gedruckter kritischer Apparat ist, kann man aber alle lesen und alle Apparate gleichermaßen ignorieren. Eine PC-gestützte, auf einer CD-ROM ausgelieferte Edition setzt jedoch voraus, dass man die für diese eine spezielle Edition benötigte Software installiert und zu benutzen lernt. Was sich bei der nächsten PC-gestützten, auf einer CD-ROM ausgelieferten Edition wiederholt, da sich weder die Software noch das realisierte Userinterface ähneln.

## 1.6 | Das WWW als einheitliches Interface

Dies änderte sich dadurch, dass sich mit der Verbreitung des WWW eine völlig neue **Umwelt für die Benutzung der Informationstechnologie** in den Geisteswissenschaften ergab. Das Interesse der bereits beschriebenen Mehrheit der Geisteswissenschaftler, die die methodischen Implikationen der technischen Verfahren nie wirklich interessiert hatten, wandten sich dem neuen Medium zu. Diejenigen, die an den Verfahren grundsätzliches Interesse zeigten, waren dagegen nach wie vor an den eigentlichen Forschungsfragen zugewandt. Das spiegelt sich deutlich darin wieder, dass das fachliche Interesse an den Jahrestagungen der Fachverbände zwischen 1995 und 2005 zeitweise sehr gering war, während zwischen 1996 und 2005 eine Tagungsserie mit dem Titel »Digital Resources for the Humanities« in Großbritannien erhebliches Aufsehen einer sehr viel weiteren Fachöffentlichkeit erreichte. Einer Fachöffentlichkeit, die darunter wohl implizit die Frage verstand, was das WWW für die Geisteswissenschaftler tun könne.

Die offizielle Beirundung für diese neue Konferenzserie las sich natürlich etwas

anders: »The overarching rationale for creating DRH was a perceived need for a broader, inclusive treatment of the creation and use of digital resources within the humanities« (<http://www.ahds.ac.uk/news/newsletters/spring-2005/spring-2005-newsletter.pdf>). Mit anderen Worten: Die in der traditionellen interdisziplinären Community als Nebeneffekt entwickelte, in der TEI ausgedruckte Vorstellung, dass es notwendig sei, projektübergreifende Ressourcen anzubieten, schob sich durch die Entstehung einer inhaltsneutralen Informationsinfrastruktur, eben dem WWW, soweit in den Vordergrund, dass die Frage der analytischen Nutzung plötzlich sekundär zu sein schien, was neue Strukturen erforderte. Eine Entwicklung, die noch dadurch verstärkt wurde, dass die Gedächtnisrichtungen – **Archive, Bibliotheken, Museen** – die innerhalb der geisteswissenschaftlichen Auseinandersetzung mit den Informationstechnologischen Möglichkeiten bisher eine sehr marginale Rolle gespielt hatten, als Herrscher über die zu digitalisierenden Bestände der **kulturellen Überlieferung** plötzlich sehr zentral wurden.

Der Namenswechsel vom traditionellen Humanities Computing zu den »Digital Humanities«, die mit dem Compendium von Schreibman, Siemens und Unsworth (2004) ins Leben traten, spiegelt letzten Endes hauptsächlich den Versuch wieder, zwei Stränge zu versöhnen: Die analytischen Traditionen, die aus den Disziplinen der Geisteswissenschaften herrühren, und versucht haben, ausgehend von einzelnen Projekten zu gemeinsamen Konzepten zu gelangen und die Vorstellung, das WWW erfordere Ansätze zur Bereitstellung von Ressourcen, die die analytische Arbeit durch andere erst ermöglichen sollten.

Zehn Jahre später können wir konstatieren, dass der Versuch eindeutig gelungen zu sein scheint: Die analytischen Traditionen des Umgangs mit Texten auf philologischer Basis, der statistischen Untersuchung geisteswissenschaftlicher Sachverhalte und vieler anderer, hier nur gestreifter, oder überhaupt nicht angesprochener, sind mittlerweile in die Diskussion um die dazu nötigen **Infrastrukturen** eng eingebunden. Was sich vielleicht am besten dadurch illustrieren lässt, dass das Paradigma digitaler Editionen, essentiell funktional vollständig vor fünfzehn Jahren und danach zunehmend geschwächt, in den letzten fünf Jahren zu einer rasant steigenden Zahl derartiger Editionen im Internet geführt hat – in einem Medium, in dem die Unterschiede zwischen den einzelnen Editionen (fast) so belanglos werden, wie zuvor im Druck.

## 1.7 | Das Beste kommt erst noch

Die Geschichte endet nicht im Jahr 2016. Auch nicht die der Digital Humanities. Zu glauben, dass die Informationstechnologie keine weitere Entwicklung haben wird, ist, wie alles Naive, amüsant. Aber auch die neuen, heute noch nicht absehbaren Entwicklungen werden daraufhin untersucht werden müssen, wieweit sie geeignet sind, als Werkzeuge den Zeitauwand für die Bewältigung von Routineaufgaben signifikant verringern können und wieweit sie neuen methodisch völlig neuen Zugangsmöglichkeiten. Dies zu untersuchen, bleibt die Aufgabe der Digital Humanities – ganz unabhängig davon, wie sie im Jahr 2066 heißen mögen.

## Literatur

- Anderson, Ian: »History and computing« (2008). [http://www.history.ac.uk/makinghistory/resources/articles/history\\_and\\_computing.html](http://www.history.ac.uk/makinghistory/resources/articles/history_and_computing.html) (26.4.2016).
- Busa, Roberto/Jones, Steven E.: »The Annals of Humanities Computing: the Index Thomisticus«. In: *Computers and the Humanities* 14 (1980), 83–90.
- Laflin, Susan: »Computer Applications in Archaeology 1973–1995« (2014). <https://www.lulu.com/10.6.2016>.
- McGann, Jerome: *Radiant Textuality: Literature after the World Wide Web*. Houndmills 2001.
- Schreibman, Susan/Siemens, Ray/Unsworth, John (Hrsg.): *A Companion to Digital Humanities*. Oxford 2004.
- Terras, Melissa/Nyhan, Julianne/Vanhoutte, Edward: *Defining Digital Humanities*. Farnham 2013.
- Winter, Thomas Nelson: »Roberto Busa, S. J., and the invention of the Machine-Generated Concordance« (1999). <http://digitalcommons.unl.edu/classicstacpub/70> (26.4.2016).
- Zweig, Benjamin: »Forgotten Genealogies: Brief Reflections on the History of Digital Art History«. In: *International Journal of Digital Art History* 1 (2015). <http://nbn-resolving.de/urn:nbn:de:bsz:16-dah-216330> (26.4.2016).

*Manfred Thaller*

## 2 Digital Humanities als Wissenschaft

### 2.1 | Die Digital Humanities: Ein weites Feld

- Verstehen wir die Digital Humanities als die Summe aller Versuche, die Informations-techniken auf den Gegenstandsbereich der Geisteswissenschaften anzuwenden, ergibt sich ein ungemäin breites Bild (ohne Anspruch auf Vollständigkeit):
- Viele Anwendungen beschäftigen sich mit **Text** als solchem. Dies kann sich auf, meist den Literaturwissenschaften dienende, Fragen nach der Entwicklung des Wortschatzes einzelner Verfasser beziehen, im Bereich der Editionsphilologie auf die beste Art, verwirrend überlieferte Texte zu rekonstruieren oder auf die Untersuchung sprachlicher Eigenschaften im Sinne der Linguistik.
  - Eine andere, große Gruppe von Projekten beschäftigt sich mit der Handhabung nicht-textueller Medien: Von Bildern, Plastiken oder Gebäudemodellen für die Kunstgeschichte und Archäologie, von Klängen und Notensätzen zum besseren Verständnis musikalischer Entwicklungen, oder von Filmen zur Beantwortung medienwissenschaftlicher Fragestellungen.
  - Wichtig, vor allem in den historischen Disziplinen, sind ferner Anwendungen, die Texte und Bilder nicht um ihrer selbst willen, sondern als Quellen von Belegen für historische Entwicklungen verwenden, sei es als Inhalte von Datenbanken, sei es als Ausgangspunkt der Untersuchung geographischer Verteilungen im Rahmen geographischer Informationssysteme.
  - Schließlich gibt es Projekte, die die neuen Technologien selbst zum Gegenstand machen: Sei es in der Analyse der Auswirkungen moderner Medien auf unser Gesellschaft, sei es zum grundsätzlichen Studium der für die Geisteswissenschaften gebotenen methodischen Möglichkeiten – bis hin zur Frage, wie geisteswissenschaftliche Forschung durch Techniken der künstlichen Intelligenz stimuliert werden kann.

### 2.2 | Die Digital Humanities: Werkzeug oder Methode?

Trotz dieser Breite des Forschungsfeldes gibt es Gemeinsamkeiten der Spielarten der Digital Humanities. Schon Padre Busa, mit dem die Geschichte der Anwendungen der Informationstechnik in den Geisteswissenschaften üblicherweise eröffnet wird (s. Kap. 1), war zunächst aus Gründen der Arbeitsökonomie an den neuen Technologien interessiert. Er wird aber gleichzeitig von vielen Forschern auch als einer der Begründer der Computeringuistik gesehen, also einer Form der Linguistik, die sich von deren anderen Ausprägungen nicht nur durch ihre **Arbeitstechniken und Werkzeuge**, sondern auch durch ihre **Fragestellungen und Methoden** unterscheidet. Dieser Gegensatz kann auch als Frage formuliert werden: Wenn Geisteswissenschaftler die Informationstechnologie einsetzen, tun sie das, um dieselben Ergebnisse, die sie auch ohne diese Technologie erreicht hätten, schneller und effektiver zu erreichen – oder streben sie Ergebnisse an, die ohne sie nicht erreichbar gewesen wären? Wenn Ergebnisse angestrebt werden, die ohne Informationstechnologie nicht erzielt werden können: Sind diese dann »besser« als die in den Disziplinen auf traditionelle Weise erzielten?

Die erste Schwierigkeit: möchte man diese Frage beantworten, ergibt sich, sobald

man die auf den ersten Blick plausible Unterscheidung zwischen Werkzeugen und Methoden zu präzisieren versucht. Durchsucht man mit technischen Mitteln einen Text von tausend Seiten nach dem Kontext, innerhalb dessen ein Wort verwendet wird, erledigt man dies etwas schneller; als es menschlichen Forschern, die dies mit dem die Zeile entlanglaufenden Bleistift ohne technische Unterstützung tun, möglich wäre. Tut man dasselbe Informationstechnisch unterstützt in einer der großen digitalen Textsammlungen, die tausende Bände enthalten, erhält man Ergebnisse, die ohne diese Unterstützung nicht erzielt werden könnten. Dadurch werden Fragestellungen möglich, die ohne das Werkzeug nicht aufkommen können – die aber in weiterer Folge auch neue **methodische Überlegungen** erfordern können.

Das beantwortet natürlich *nicht* die Frage, ob Geisteswissenschaftler sich mit Problemen beschäftigen *sollen*, die nur beantwortet werden können, wenn man tausende von Seiten durchsucht. Man kann auch der Ansicht sein, dass sich die Geisteswissenschaften ganz bewusst auf Fragestellungen konzentrieren sollen, die innerhalb eines kleinen Bestandes an herausragenden Texten – beispielsweise denen einzelner Verfasser – beantwortet werden können. Die Gegenpositionen sind etwa folgende: Der Auffassung, dass die Geisteswissenschaft lehren soll, die gedanklichen (und zuvor: sprachlichen) Strukturen einzelner Verfasser möglichst detailliert zu schätzen und zu verstehen, kann eine andere gegenüberstehen, die betont, dass es wichtig sei, dieselben Strukturen innerhalb des gedanklichen (und zuvor: sprachlichen) Kontexts des Jahrhunderts der jeweiligen Verfasser zu verstehen. Und daraus kann schließlich die Vorstellung erwachsen, dass eigentlich das Verständnis der aus dem sprachlichen Befund ableitbaren gedanklichen Strukturen einer Epoche wichtiger sei, als die einzelner Verfasser.

### 2.3 | Die Digital Humanities im Kontext der geisteswissenschaftlichen Disziplinen

Die Entscheidung, welche dieser drei Fragestellungen – Verständnis einzelner Verfasser, Verständnis des Verhältnisses einzelner Verfasser zu den intellektuellen Kontexten, in denen sie lebten, Verständnis dieser intellektuellen Kontexte – am wichtigsten sei, ist selbstverständlich eine, die nicht aus den Digital Humanities allgemein hergeleitet werden kann, sondern aus denen der inhaltlichen Disziplinen abgeleitet werden muss. Welche Werkzeuge der Informationstechnologie für welche geisteswissenschaftlichen Fächer besonders angemessen sind, hängt ganz wesentlich vom **Selbstverständnis des jeweiligen Faches** ab. So gab es in den 1960er und 1970er Jahren starke Strömungen in den Geisteswissenschaften, Fragestellungen und Methoden der Sozialwissenschaften aufzugreifen. Ganz besonders in den Geschichtswissenschaften, aber auch in der ganzen Breite der geisteswissenschaftlichen Fächer führte das zu einem großen Interesse an quantitativ/statistischen Verfahren beziehungsweise an der Nutzung von Datenbanken. Die Nutzung beider Techniken ging zurück, als das interdisziplinäre Interesse an sozialwissenschaftlichen Ansätzen und damit mittelbar an deren empirischen Vorgehensweisen wieder abnahm.

Das gilt ganz explizit nicht nur den Arbeitstechniken, sondern auch den methodischen Annahmen, die ihre Anwendung erfordern oder mindestens plausibel machen. In erster Linie ist hier die Einstellung gegenüber dem **Verhältnis zwischen den Geistes- und den Naturwissenschaften** zu verstehen. Im Deutschen – und der deutschen akademischen Community – besteht überwiegend die Vorstellung, dass

es eine Vorstellung von Rationalität gäbe, die allen ›Wissenschaften‹ gemein sei. Auch wenn es, nach Forschern wie Wilhelm Dilthey, fundamentale, methodische Unterschiede gibt, sind Geistes- und Naturwissenschaften deshalb trotzdem nicht wesentlich voneinander getrennt. Im Englischen gibt es das Konzept ›Wissenschaft als übergeordneter Begriff‹, der das Gegensatzpaar ›Science‹ und ›Humanities‹ vereint, nicht; daraus entsteht dann leicht die Vorstellung, dass es zwischen Naturwissenschaften und Geisteswissenschaften keine Berührungspunkte gäbe, ja, dass die Geisteswissenschaften der Kunst sehr viel näher stünden als der ›Wissenschaft‹.

Unabhängig von dieser Frage, derer man sich beim Lesen englischsprachiger Literatur zu den Methoden der Digital Humanities stets bewusst sein sollte, hängt die Einstellung zu methodischen Folgen aus der Anwendung informationstechnischer Werkzeuge ganz besonders davon ab, welche Position Forscher in der Frage haben, was eine **angemessene Beweisführung** in ihrer geisteswissenschaftlichen Disziplin sei. Wird mit Paul Feyereabend überhaupt verneint, dass es so etwas wie eine verbindliche Methode geben könne, so beschränkt sich die Relevanz der Digital Humanities darauf, Werkzeuge für den Umgang mit Dokumenten und Bildern auf rein technischer Ebene bereitzustellen. Wird die Bedeutung der **Hermeneutik**, im Sinne holistischer Durchdringung eines Sachverhaltes durch ein individualistisches Verständnis betont, können die Digital Humanities aber bereits methodisch signifikant werden, weil der Umgang mit digital gespeicherter Information oft leichter nachvollziehbar und für andere dokumentierbar ist, als dies bei der Verwendung persönlicher Notizen der Fall wäre. Bei allen Ansätzen schließlich, bei denen ein **Beweis** im Sinne quantitativ/statistischer Argumentation verlangt wird, ist der methodische Schwerpunkt eindeutig.

Zu betonen ist dabei jedoch, dass selbst Paul Feyereabend Werkzeuge der Digital Humanities hätte verwenden können, um ihn interessierende Texte zu finden: Die Digital Humanities insgesamt erlauben also methodische Positionen, verlangen sie aber nicht und finden in der ganzen Breite der möglichen methodischen Haltungen der Geisteswissenschaften ihre Berechtigung. Die Digital Humanities sind **methodennutral**, trotz der methodischen Implikationen, die wir einleitend erwähnt haben.

### 2.4 | Die Digital Humanities jenseits einzelner geisteswissenschaftlicher Disziplinen

Unabhängig davon, wie man sie gegenüber den Naturwissenschaften abgrenzt, besteht jedenfalls Konsens, dass es von ihnen unterschiedene Geisteswissenschaften gibt. Trotzdem kann man kein Studium der *Geisteswissenschaft* beginnen, sondern nur eines einer spezifischen Disziplin: Germanistische oder Historische Seminare befragen häufig an den Universtitäten, ›geisteswissenschaftliche‹ nie. Wieweit kann es dann eine digitale Geisteswissenschaft als universitäres Fach geben, wenn es keine Geisteswissenschaft als solche gibt?

Auf der Werkzeugebene gibt es eine einfache Antwort: Für Geisteswissenschaftler, die sich dafür interessieren, wie man möglichst effektiv in großen Mengen an Texten sucht, ist es zunächst unerheblich, ob sie dies tun, weil sie feststellen wollen, ob eine sprachliche Konstruktion für bestimmte Verfasser typisch ist, oder weil sie wissen wollen, ab wann die Schlagworte einer bestimmten politischen Richtung in digitalen Zeitungen auftauchen. Das Handwerkszeug ist für die philologischen Disziplinen das gleiche, wie für die der historischen Forschung. Daran könnte man so-



gar eher die Frage anschlüssen, ob diese Neutralität des Werkzeugs nicht so ausgeprägt sei, dass es auch unethisch werde, ob Geisteswissenschaftler oder Physiker Texte durchsuchen. Auf allgemeiner technischer Ebene wäre das auch zu bejahen: Der kompetente Umgang mit den informationstechnischen Werkzeugen, die heute allgemein verbreitet sind, ist für alle wissenschaftlichen Disziplinen nötig und begründet keine akademische oder fachspezifische Qualifikation.

Intuitiv kann man den Unterschied zwischen **allgemeiner Kulturtechnik** und geisteswissenschaftlicher, aber **fachübergreifender, Qualifikation** dort sehen, wo die Anwendung auf die fachspezifischen Daten oder Texte fachspezifische Kenntnisse voraussetzt. Nach wie vor intuitiv wird dies illustriert durch die Probleme, die sich ergeben, wenn wir große Textmengen aus dem 18. Jahrhundert bearbeiten wollen: Auch hier sind die nötigen Qualifikationen zum Umgang mit Texten in älterer Orthographie und älteren Sprachformen unabhängig davon, ob hinter der Beschäftigung mit dem Material eine philologische oder eine historische Fragestellung steht. Für beide geisteswissenschaftlichen Hintergründe sind diese Qualifikationen notwendig. Für die Beschäftigung von Naturwissenschaftlern mit gegenwartsbezogenen Texten nicht.

Und diese disziplinübergreifenden Eigentümlichkeiten gehen über den Umgang mit Texten hinaus. Sollen räumliche Phänomene in historische Landkarten eingezeichnet werden, ist es unabhängig vom disziplinären Interesse notwendig zu verstehen, wie sich die Methoden, nach denen diese Karten gezeichnet wurden (die sogenannte »Projektion«), von denen für heutige Karten üblichen unterscheiden (s. Kap 21.2). Werden Noten zur Klärung musikwissenschaftlicher Fragen analysiert, ist zu klären, wie man die vorgefundene Notation verarbeitet – soll eben komponierte Musik in Notenform ausgegeben werden, kann einfach jenes Verfahren gewählt werden, das das zur Eingabe der Noten verwendete Programm vorgibt.

Diese Notwendigkeit, aus den Quellen, die zur Untersuchung des die Geisteswissenschaftler interessierenden Phänomens vorliegen, Daten zu extrahieren und in eine Form zu bringen, die ihre Untersuchung mit informationstechnischen Mitteln erlaubt, ist Bestandteil des Konzepts der **Modellierung** (s. Kap. 7). Hier liegt der fachübergreifende Kern der Digital Humanities: Es besteht Konsens darüber, dass für die adäquate Modellierung von geisteswissenschaftlichen Sachverhalten für die Informationstechnik Kenntnisse und Vorgehensweisen benötigt werden, die mehreren geisteswissenschaftlichen Disziplinen gemein sind, sich aber von den in den Naturwissenschaften benötigten unterscheiden. Um einem Missverständnis vorzubeugen: Die Modellierung von Sachverhalten in einer Form, die sie für informationstechnische Systeme verarbeitbar macht, ist eine unabhängige Voraussetzung für den Einsatz solcher Systeme in absolut jeder Anwendungsdomäne. Die dafür einzusetzenden Konzepte in den Geisteswissenschaften unterscheiden sich jedoch von denen in anderen Disziplinen signifikant. Um einem anderen Missverständnis vorzubeugen: Modelle an sich sind den Geisteswissenschaften nicht fremd. Die für den Einsatz der Informationstechnologien notwendigen stehen in einer langen Tradition, die vor allem Willard McCarty sehr plausibel beschrieben hat. Er lässt seinen **Stammbaum der Modellierung** bei der Verwendung von veranschaulichenden **Analogen** beginnen, die versuchen, den Wesenskern eines Problems sichtbar zu machen und führt ihn über verschiedene Darstellungstechniken weiter zu Konzepten wie dem Experiment und der Simulation.

Vertrautes und Neues stehen hier nebeneinander: Dass die Zusammenhänge zwischen *Knowledge* durch ein *Diagramm* anschaulicher werden, ist ein allgemein be-

kanntes Beispiel für ein Modell, in diesem allgemeineren Sinn. Simulationen – etwa die im Fernsehen gezeigten visuellen Simulationen der Erde umkreisender Satelliten – scheinen den Geisteswissenschaften zunächst wesensfremd. Denken wir an Beispiele, wo die archaischen Fächer 3D-Modelle generieren, die es erlauben, sich, unter Umständen durch Techniken der Virtualen Realität, in virtuellen Rekonstruktionen nicht mehr existierender Gebäude zu bewegen, haben wir jedoch ein unmittelbar plausibles Beispiel für diese Art von »Modell« in den Geisteswissenschaften, das offenbar eine sehr viel detailliertere »Modellierung« erfordert, als die in ein Diagramm mündende. Dass solche **Simulationen** weit über die Wiedergabe nicht mehr existierender Objekte hinausgehen und sich auch auf die Untersuchung rein konzeptueller Zusammenhänge auswirken können, sei hier nur angemerkt. Beispiele für die Praxis des »Modellierens« finden sich in anderen Kapiteln dieses Bandes.

## 2.5 | Die Digital Humanities und die Informatik

Ein 3D-Modell, das ein zerstörtes Gebäude im virtuellen Raum wieder erstehen lässt, liegt im Jahr 2016 immer noch ein wenig nach Hightech. Die Fähigkeit, in einigen tausend Büchern nach Regelmäßigkeiten zu suchen, die als Kontext der Beobachtung an einem einzelnen Text verwendet werden können, eigentlich nicht minder, **ist** das, die Digital Humanities hätten ein besonders enges Verhältnis zur Informatik.

Nicht unbedingt, noch nicht einmal häufig. »Die Informatik« umfasst mittlerweile ein Spektrum fachlicher Spezialisierungen, das kaum weniger breit ist, als das der Geisteswissenschaften, ganz sicher nicht weniger breit, als das einer ihrer großen Teilbereiche, etwa »der Philologien«. Der Kern der Informatik besteht stets in der Untersuchung der Frage, wie man Information am Rechner darstellen und verarbeiten kann. Nun haben wir zuletzt eben auf die Verwendung von 3D-Modellen verwiesen: **Was** die Frage, wie man ein solches, zunächst sehr komplexes Modell auf dem Rechner darstellt, dann ein Beispiel dafür sein, dass sich Informatik und Geisteswissenschaften, hier repräsentiert durch die Archäologie, eng berühren? Nicht notwendiger Weise. Zwar gibt es innerhalb der Informatik in der Tat ein Spezialgebiet, das sich mit genau dieser Frage – wie man 3D-Modelle definiert und visualisiert – auseinandersetzt, in aller Regel findet diese Auseinandersetzung aber auf einem sehr viel allgemeineren Niveau statt, wo man sich nicht mit der Frage beschäftigt, wie **man** ein ganz bestimmtes Gebäude modelliert und darstellt, sondern wie man diese **Frage** für **alle überhaupt vorstellbaren Objekte** lösen kann. Was mutatis mutandis kann noch wieder für den Umgang mit großen Textmengen gilt.

Während es einzelne Ansätze zur Untersuchung der Eignung von Grundkonzepten der Informatik auf die Geisteswissenschaften gibt, beschäftigen sich nahezu alle Bereiche der Digital Humanities jedoch mit der Frage, wie man die Informationstechnologie, also die aus den abstrakten Konzepten der Informatik abgeleiteten **praktischen Lösungen**, für die Geisteswissenschaften nutzbar machen kann.

## 2.6 | Die Digital Humanities

Beispiele der Digital Humanities aus dem Bereich der Quantitativen Analyse (s. Kap. 20) haben gemein, dass die quantitative Aussage letzten Endes der detaillierten Betrachtung der Texte folgt. Die Beschäftigung mit dem intensiv gelesenen Text steht also am Anfang, die quantitative Betrachtung folgt. Wir gingen in diesem Kapitel schon darauf ein, dass es Fragen geben könne, denen wir uns nähern könnten, wenn wir eine Sammlung von tausenden von Bänden durchsuchen könnten. Wenn wir tausende von Bänden nach Wörtern durchsuchen können, können wir diese Wörter auch zählen. Wir können diese tausende von Bänden aber nicht mehr intensiv lesen. Die quantitative Ergänzung der verstehenden Interpretation wird also nicht mehr statistisch ergänzt (**deskriptive Statistik**) oder methodisch abgesichert (**schließende Statistik**), sondern sie fehlt zunächst, und die statistische Annäherung an die große Textmenge ist Ausgangspunkt der Analyse (und kann ggf. erst sekundär durch die intensive Lektüre von einzelnen Texten im quantitativ erschlossenen Gesamtbestand ergänzt werden).

In den Diskussionen der letzten Jahre wird der Zugang zu Texten über die intensive Lektüre einzelner Texte (der durch statistische Analysen ergänzt werden kann, aber nicht muss) oft als **close reading** bezeichnet; der Zugang zu großen Textsammlungen über statistische Aussagen über die Texte in der Sammlung insgesamt (der durch die intensive Lektüre von Einzeltexten ergänzt werden kann, aber nicht muss) als **distant reading**.

Wir haben damit begonnen aufzuzeigen, dass sich hinter diesem Begriff Anwendungen verbergen, die scheinbar wenig gemeinsam haben: Was hat die Rekonstruktion des ursprünglichen Bauzustandes eines Barockschlosses in der Virtualen Realität mit der statistischen Analyse frühneuzeitlicher Steuerlisten zu tun?

Was sie miteinander und dem gewählten Beispiel zu den quantitativen Verfahren gemein haben, ist zweierlei. Sie bieten Werkzeuge an, mit denen die Forschungsgegenstände der Geisteswissenschaften, mit unveränderter Methodologie, intensiver und gründlicher untersucht werden können. Und sie *können* wissenschaftstheoretische Fragen eröffnen. Aristotelisch ausgedrückt: Wenn wir *können* wissenschaftstheoretischen Fragen in einem Jahrhundert gedruckten literarischen Werken haben, ist die Konzentration auf einen Kanon von hundert besonders wichtigen ein Teil der **Substanz** der geisteswissenschaftlichen Arbeit? Oder ist es ein **Akzidens**, zufälliges Nebenprodukt der Arbeitsbedingungen unter denen sich die Geisteswissenschaften vor der digitalen Verfügbarkeit der großen Textmengen entwickelt haben?

### Literatur

- Bodenhamer, David J./Corrigan, John/Harris, Trevor M. (Hg.): *The Spatial Humanities*. Bloomington/Indianapolis 2010.
- Gold, Matthew K. (Hg.): *Debates in the Digital Humanities*. Minneapolis/London 2012.
- McCarty, Willard: *Humanities Computing*. Houndmills 2005.
- Moretti, Franco: *Distant Reading*. London 2013.
- Schreibman, Susan/Siemens, Ray/Unsworth, John (Hg.): *A New Companion to Digital Humanities*. Chichester 2016.
- Thaller, Manfred (Hg.): »Controversies around the Digital Humanities«. In: *Historical Social Research* 37/3 (2012), 7–23.