

## Daten und ihre Basis: Chancen und Probleme beim Umgang mit originär digitalen Quellen

Im Jahr 2011 formulierte der Zeithistoriker Kiran Klaus Patel die Frage "wer archiviert eigentlich die SMS von Angela Merkel?"<sup>1</sup> Obwohl diese Frage bereits zwei Jahre später überraschend von Edward Snowden beantwortet wurde, bleiben Patels Appelle an die Geschichtswissenschaften aktuell: Patel verweist auf die Relevanz digitaler Kommunikation zum Verständnis und der Erforschung der jüngsten Vergangenheit, auf die Dringlichkeit, dieser Relevanz in der archivalischen Sammlungs- und Bewahrungspraxis zu entsprechen und appelliert schließlich (implizit) an die eigene Zukunft, sich auf die methodischen Herausforderungen im Umgang mit einer scheinbar neuen Quellengattung einzulassen.

Heute, 10 Jahre später, ist die Digital History aus den Diskussionen des Faches nicht mehr wegzudenken, auch wenn noch keine allgemein akzeptierte Festlegung ihrer Methoden und Gegenstände verfügbar ist – und vielleicht auch gar nicht wünschenswert. Auffällig ist jedoch, dass Arbeiten zur Zeitgeschichte und speziell solche, die dezidiert originär digitale Quellen untersuchen, nach wie vor eher rar sind. Dies ist keinesfalls nur einem methodischen Konservatismus der Disziplin geschuldet, sondern hat auch mit dem Charakter digitaler Quellen zu tun. Ihre Erforschung stellt die Geschichtswissenschaft vor erhebliche und bislang ungelöste methodische und analytische Probleme, die im Folgenden anhand von drei Paradoxien vorgestellt werden. Dem schließt sich eine kurze Skizze vielversprechender theoretisch-konzeptioneller Zugänge zu originär digitaler Quellen (digi-born) an, deren Potential darauf am Beispiel eines Datenbanksystems aus dem Vietnamkrieg illustriert wird.

I. Drei Paradoxa: Überfluss und Mangel, Stabilität und Volatilität, Virtualität und Materialität

Der Digital History mangelt es nicht an Manifesten und euphorischen Programmatiken, in denen ihre Innovationskraft, die durch sie ermöglichte Erweiterung der Perspektiven und die breitere empirische Fundierung geschichtswissenschaftlicher Forschung durch die von ihr zur Verfügung gestell-

---

<sup>1</sup> Kiran Klaus Patel, 'Zeitgeschichte im digitalen Zeitalter. Neue und alte Herausforderungen' in: *Vierteljahresshefte für Zeitgeschichte* 59/3 (2011), S. 331-351.

ten Methoden gefeiert und beschworen wird.<sup>2</sup> Wenngleich teilweise als Repliken gegen die Vorbehalte und Herablassungen der etablierten historischen Disziplinen formuliert, so sind diese Texte doch von der Gewissheit getragen, dass die Grundtendenz zur Fortschreitenden Digitalisierung auch zu einer quantitativen Steigerung der verfügbaren Quellen führen würde. Angesichts der seit der Jahrtausendwende laufenden großen Projekte der Retro-Digitalisierung durch Archive und Bibliotheken sowie des erleichterten Online-Zugangs zu diesen Beständen, entspricht die oft formulierte Wahrnehmung eines Überflusses an möglichen Quellen durchaus der Alltagserfahrung praktizierender Historikerinnen und Historiker.<sup>3</sup> Das gleichzeitige Anwachsen digitaler Kommunikationsmedien im Internet und der exponentiellen Vermehrung des allgegenwärtigen Arsenal an Geräten und Praktiken der Informationstechnologie, ließ die Annahme gerechtfertigt erscheinen, dass die Tendenz zu einem steten "mehr" an Quellen durch den enormen, scheinbar allgemeinen und globalen Ausstoß von digitalen Spuren und Zeugnissen noch einmal erheblich verstärken müsse.

In krassem Gegensatz stehen diesen Erwartungen zu Warnungen vor einem informationellen "schwarzen Loch" digitaler Amnesie, ja, einem "dunklen Zeitalter".<sup>4</sup> Die Archivwissenschaft befasst sich bereits seit den 1990er Jahren intensiv mit dieser Problematik, wohl aus der praktischen Erfahrung, dass in dieser Zeit digitale Dokumente meist ausgedruckt an Archive übergeben wurden (und damit als interpoliert und selektiv gelten konnten), oder dass Archive digitale Quellen und Artefakte stetig in neue Formate überführen mussten, was ebenfalls einer Interpolation gleichkommen konnte und einige weitere Fragen der Langzeitbewahrung aufwarf.<sup>5</sup> Die rapide technische Obsoleszenz digitaler Speichermedien und Plattformen bei der gleichzeitig an Bedeutung gewinnenden Frage der Selektion der zu bewahrenden Zeugnisse angesichts einer steigenden Menge von digitalen Hinterlassenschaften beschäftigt nicht nur das

---

<sup>2</sup> Jo Guldi und David Armitage, *The history manifesto* (Cambridge, United Kingdom: Cambridge University Press, 2014); Shawn Graham, Ian Milligan, und Scott Weingart, *Exploring big historical data: the historian's macroscope* (London: Imperial College Press, 2016).

<sup>3</sup> „Scarcity or Abundance? Preserving the Past in a Digital Era – Roy Rosenzweig Center for History and New Media“, URL: <http://rrchnm.org/essay/scarcity-or-abundance-preserving-the-past-in-a-digital-era/>; Ian Milligan, *History in the Age of Abundance?: How the Web Is Transforming Historical Research* (Montreal: McGill-Queen's University Press, 2019).

<sup>4</sup> Oliver Kichle, Digitalia fundamentalis – Speicher und Zugriff, in: Diskrete Werte (blog), URL: <https://digitalia.hypotheses.org/21>.

<sup>5</sup> Einer der frühesten und nach wie vor umtriebigen Mahner vor einem Verlust des digitalen Kulturerbes ist der ehemalige RAND-Forscher Jeff Rothenberg, siehe: Avoiding Technological Quicksand: Sections 1-3, URL: <https://www.clir.org/pubs/reports/rothenberg/introduction/#longevity>; Rothenberg, Preserving Authentic Digital Information; URL: <https://www.clir.org/pubs/reports/pub92/rothenberg/>.

wissenschaftliche Personal von Archiven und Bibliotheken, sondern auch naturwissenschaftliche Forschungszentren<sup>6</sup> und Privatpersonen.<sup>7</sup>

Archivwissenschaften und Geschichtswissenschaften haben sich dieser Problemlage durch den Versuch der gemeinsamen Entwicklung von Standards zur Bewahrung und des Zugangs gewidmet.<sup>8</sup> Auch die Zugänglichkeit zu archivierten originär digitalen Quellen und vor allem ihre Anreicherung mit Metadaten und deren Archiv-übergreifende Verknüpfung (Linked Open Data, LOD), wird von interdisziplinären Diskussionsgruppen und Initiativen – wie etwa der Konsortialinitiative 4Memory – mit Leidenschaft vorangetrieben.<sup>9</sup> Letztlich aber verweisen sämtliche solcher Anstrengungen zur Etablierung einheitlicher Standards in der Sammlung, Bewahrung und Qualitätssicherung originär digitaler Quellen auf die enormen und dieser Quellenform inhärenten Anforderungen: Das exponentielle quantitative Wachstum möglicher Quellen; der damit einhergehende Mehraufwand in der Anreicherung und Beschreibung der Bestände mit entsprechenden Metadaten; und nicht zuletzt die technischen, personellen und auch energetischen Schwierigkeiten, der Obsoleszenz historischer Systeme und Medien langfristig und nachhaltig zu begegnen.<sup>10</sup> Originär digitale Quellen

---

<sup>6</sup> <https://cerncourier.com/a/data-preservation-is-a-journey/>

<sup>7</sup> <https://www.theguardian.com/technology/2015/feb/13/google-boss-warns-forgotten-century-email-photos-vint-cerf>

<sup>8</sup> Der älteste und am weitesten verbreitete Standard, das seit 1997 erarbeitete Open Archival Information System (OAIS) Referenzmodell, legte den Grundstein für eine einheitliche Bewahrungs- und Beschreibungspraxis, und hat seit seiner Einführung 2002 zahlreiche Erweiterungen und Ergänzungen erfahren. Ein Überblick findet sich bei: Erin Baucom, 'A Brief History of Digital Preservation', in: Mansfield Library Faculty Publications 31 (2019), URL: [https://scholarworks.umt.edu/cgi/viewcontent.cgi?article=1033&context=ml\\_pubs](https://scholarworks.umt.edu/cgi/viewcontent.cgi?article=1033&context=ml_pubs); sowie Thorsten Ries und Gábor Palkó, „Born-Digital Archives“, *International Journal of Digital Humanities* 1, Nr. 1 (April 2019), S. 1–11, <https://doi.org/10.1007/s42803-019-00011-x>.

<sup>9</sup> Frank Kell, 'LINKAGE — Digitale Gegenwart und Zukunft Historischer Forschung. Die Ziele der Konsortialinitiative 4Memory', URL: <https://blog.historikerverband.de/2020/11/11/linkage-digitale-gegenwart-und-zukunft-historischer-forschung/>.

<sup>10</sup> Aus archivalischer Perspektive, siehe: Michael Friedewald und Timo Leimbach, Computersoftware als digitales Erbe. Probleme aus der Sicht der Technikgeschichte. In: Robertson-von Trotha, Hauser, Caroline u. Robert (Hg.): Neues Erbe. Aspekte, Perspektiven und Konsequenzen der digitalen Überlieferung. Karlsruhe 2011, S. 201-219; Siegrid Schieber, Digitale Archivierung im Hessischen Landesarchiv, *ABF Technik*, 37,2 (2017), 93-102; URL: <https://www.degruyter.com/view/journals/abitech/37/2/article-p93.xml>; doi: <https://doi.org/10.1515/abitech-2017-0022>; Andrea Hänger, Die Mühsal der Zugänglichkeit. Methodische Herausforderungen beim Aufbau virtueller Lesesäle, in: *Zeithistorische Forschungen/Studies in Contemporary History*, Online-Ausgabe, 17 (2020), H. 1, URL: <https://zeithistorische-forschungen.de/1-2020/5828>, DOI: <https://doi.org/10.14765/zzf.dok-1763>, Druckausgabe: S. 179-189; Christian Keitel, Digitale Archivalien, in: *Südwestdeutsche Archivalienkunde*, URL: <https://www.leobw.de/themenmodul/sudwestdeutsche-archivalienkunde/querschnittsartikel/digitale-archivalien>; Gerald Maier, Die Bedeutung der Archive für Forschungsdaten in der Geschichtswissenschaft, URL: <https://blog.historikerverband.de/2020/11/11/die-bedeutung-der-archiv-fuer-forschungsdaten-in-der-geschichtswissenschaft/>.

sind also von der paradoxen Situation gekennzeichnet, dass gleichzeitig ein Überfluss und ein Mangel zu herrschen scheint.

Ein zweites Paradox, das digi-born Quellen inhärent ist, betrifft ihre gleichzeitige Flüchtigkeit (oder Volatilität) und Stabilität. Im Zusammenhang mit der bereits angeschnittenen Obsoleszenz und der andauernden Suche nach nachhaltigen Strategien der Langzeitarchivierung beschreibt etwa Paul Dourish dieses paradoxe Verhältnis:

On one hand, digitization and digital information are often associated with preservation and archival practices; think, for example, of efforts to digitize ancient manuscripts to make them available even as material on which they are written crumbles, or, on a more mundane level, of the services that promise to "preserve" your family members by digitizing old photographs, film, and videotape. Yet the digital devices on which these digital materials are dependent are not themselves eternal or even particularly long-lived. Computer-systems from even just a few years ago are already obsolete and aging. Moreover, a considerable leap of faith is required to imagine digital media as long standing: given that the very earliest general-purpose computer systems became operational only in the 1940s, there could literally be no digital storage format that has ever lasted more than seventy years.<sup>11</sup>

Diese Paradoxie setzte sich von den Medien und Formaten fort hin zu den Dateien und Daten selbst, wie wir von unserem täglichen Umgang mit Schreibprogrammen und Emails wissen: Unsere Texte und Korrespondenzen existieren in mehreren Stadien, vom Entwurf bis zur ersten publikationsfähigen Version durchlaufen sie mehrfache Durchgänge der Veränderung. Auch Websites, Chatverläufe und natürlich allen voran Datenbanken sind einer stetigen Erneuerung und Aktualisierung unterworfen, eine Eigenschaft, die der Etymologie des Begriffs "Datum" als etwas Gegebenes, festes, zuwiderläuft.<sup>12</sup> Damit haben Datenbanken und die in ihnen enthaltenen Informationen einen inhärent provisorischen und volatilen Charakter, der sowohl ihre Archivierung als auch ihre geschichtswissenschaftliche Analyse stets mit der Fragen konfrontiert, ob es sich bei einer Quelle um einen "Schnappschuss" oder um den "letzten Stand"

---

<sup>11</sup>Paul Dourish, *The stuff of bits: an essay on the materialities of information* (Cambridge, Massachusetts: The MIT Press, 2017), S. 62.

<sup>12</sup>Zum Begriff, siehe: Daniel Rosenberg, 'Daten vor Fakten', in Ramón Reichert (Hg.), *Big Data. Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie* (Bielefeld: transcript-Verlag, 2014), 133–56.

handelt, und was dies jeweils für ihre Interpretation bedeutet. Mehr noch: viele digi-born Quellen sind nicht nur dynamisch, sondern auch interaktiv. Vor allem bei Datenbanken, aber auch bei anderen originär digitalen Artefakten ist zu konstatieren, dass einzelne Datensätze und Datenpunkte überhaupt erst durch Interaktion – etwa durch Abfrage – entstehen. In der vom BKA unter Horst Herold aufgebauten Datenbank zur Suche und Bekämpfung der RAF wären demnach nicht die einzelnen Einträge, wie das Register sämtlicher amtlich gemeldeter Kfz-Kennzeichen von historischem Interesse, sondern die Strategie der Beamten bei ihrer Durchsuchung und Abfrage, die jedoch der Datenbank selbst äußerlich sind.<sup>13</sup>

Im Widerspruch zur viel diskutierten Flüchtigkeit originär digitaler Quellen steht ihre bemerkenswerte Stabilität, die sich der einfachen Kopierbarkeit und damit Loslösung vom jeweiligen Speichermedium verdankt. Die meisten digi-born Quellen existieren bereits in mehreren Kopien, noch bevor sie bewusst vervielfältigt oder gespeichert werden. Matthew Kirschenbaum diskutiert die erstaunliche Stabilität und Permanenz digital-elektronischer Dateien eingehend:

Creating a file and saving it to a hard drive does not yield a simple one-to-one correspondence between the document (or file of whatever type) and its record on the disk. First, word processors and other productivity software routinely include an auto-save function that writes a snapshot of an open file to the disk at set intervals. The presence of such files is not always obvious: often they have opaque or arbitrary-seeming names (a copy of this document, for example, currently exists in one of my temporary directories as ~WRL0005.tmp). Nor do they always appear in standard directory listings (this same file would be invisible to me if the directory was configured to only display files with common extensions). This phenomenon is sometimes known as "ambient data", the term emphasizing the way in which records accumulate on a file system absent the intervention of any single, conscious (human) agency. Most computers also use a portion of their hard disk as an extension of their RAM, a type of storage known as virtual memory or swap space.<sup>14</sup>

---

<sup>13</sup> David Gugerli und Hannes Mangold, „Betriebssysteme und Computerfahndung: Zur Genese einer digitalen Überwachungskultur“, *Geschichte und Gesellschaft* 42, Nr. 1 (2016), S. 144–74.

<sup>14</sup> Matthew G. Kirschenbaum, *Mechanisms: New Media and the Forensic Imagination*, (Cambridge, Mass: MIT Press, 2012), S. 51.

Dieser Prozess der Vervielfältigung durch identische Kopien setzt sich auf bewussterer Ebene durch Backup-Routinen klügerer Computer-Nutzerinnen und Nutzer fort sowie auf eher unbewusster Ebene durch zahllose Kopien, die eine simple Email auf ihrem Weg durch das Netz auf allen beteiligten Servern generiert. Diese Vervielfältigung und die durch sie bewirkte Stabilisierung erschöpft sich nicht im (gewissermaßen "semantischen") Inhalt von Dateien, sondern erfasst auch eine Myriade in ihnen enthaltener Metadaten wie Zeitstempel und Log-files. Originär digitale Quellen sind also von der paradoxen Qualität gekennzeichnet, dass sie gleichzeitig als ephemere, flüchtig und volatil beschreiben werden können, aber auch als stabil und persistent.

Ein drittes Paradox, das digi-born Quellen zu eigen ist, liegt in ihrer gleichzeitigen Virtualität und Materialität begründet. Diese widersprüchliche Dualität ist auch die Grundlage für die beiden vorgenannten Paradoxa. Digitale Objekte sind gekennzeichnet durch ihre Repräsentation in der universellen Grammatik der Leibniz'schen Binärzahlen. Diese Eigenschaft macht sie beliebig kopierbar und auf verschiedensten Systemen und Formaten darstell- und speicherbar. Wir alle gehen im Alltag selbstverständlich davon aus, dass sich der Inhalt unserer Manuskripte, der Sound unserer Musikdateien oder die Gegenständen unserer favorisierten Websites nicht ändern werden, ganz gleich wie und wo wir sie speichern oder von welchem Gerät wir sie abrufen. Diese geradezu metaphysische Eigenschaft digitaler Artefakte hat zu einer eigenartigen Wiederkehr des Idealismus in feuilletonistischen und geisteswissenschaftlichen Diskussionen geführt, in denen Begriffe wie Virtualität oder Ubiquität und Metaphern wie die der "cloud" eifrig aufgenommen, reproduziert und auf weitere Bereiche der politischen, sozialen und kulturellen Gegenwart übertragen wurden. Diese idealistische Wahrnehmung wird jedoch zunehmend in Frage gestellt und mitunter scharfer Kritik unterzogen durch Arbeiten, die auf die infrastrukturellen Voraussetzungen der Digitalisierung und des Internets verweisen, oder ihre ökologischen Kosten betonen und ihre Grundlagen in einer globalen sowie innergesellschaftlichen ökonomischen Ungleichheit und ausbeuterischen Arbeitsverhältnissen herausarbeiten.<sup>15</sup> Abgesehen von dieser not-

---

<sup>15</sup> Siehe beispielhaft: Lisa Parks und Nicole Starosielski (Hg.), *Signal traffic: critical studies of media infrastructures, The geopolitics of information* (Urbana: University of Illinois Press, 2015); Dirk van Laak, *Alles im Fluss: die Lebensadern unserer Gesellschaft-Geschichte und Zukunft der Infrastruktur*, Geschichte (Frankfurt am Main: S. Fischer, 2018); Michael Homberg, 'Mensch / Mikrochip. Die Globalisierung der Arbeitswelten in der Computerindustrie 1960 bis 2000 – Fragen, Perspektiven, Thesen', *Vierteljahrshefte für Zeitgeschichte* 66, Nr. 2 (2018), S. 267–93; Nathan Ensmenger, 'The Environmental History of Computing', *Technology and Culture* 59, Nr. 4S (2018): S7–33; Robert Markley, 'Bounda-

wendigen Einbettung digitaler Objekte in die physische Um-Welt wirft die Materialität digitaler Medien bedeutsame Fragen für Archiv- und Geschichtswissenschaft auf. Denn wenngleich digitale Objekte nicht notwendig an individuelle Computersysteme und Speichermedien gebunden sind (im Sinn von "ein ganz bestimmtes System oder mein persönlicher Computer/USB-Stick"), so können sie doch notwendigerweise nur in und auf spezifischen Speichern und Systemen existieren und funktionieren. Originär digitale Quellen sind wie alle digitalen Objekte stets mindestens an einem spezifischen Ort physisch und materiell präsent und funktional von einem spezifischen informationellem System zu ihrer Verarbeitung und Darstellung abhängig. Wie Matthew Kirschenbaum in seiner faszinierenden Analyse der "Grammatologie des Hard-Drives" überzeugend herausgearbeitet hat, handelt es sich auch bei digital gespeicherten Informationen letztlich um "Inschriften", also physischen und materiellen Repräsentationen auf einer Trägersubstanz, etwa in Form von elektromagnetischen Polarisierungen auf einer Festplatte, CD-ROM oder Diskette, oder in den festgehaltenen elektrischen Spannungen in den Zuständen der einzelnen Halbleiter in einem Flash-Speicher, wie etwa in USB-Sticks oder SD-Karten.<sup>16</sup> Damit unterscheiden sich die modernsten Speichermedien in ihren Funktionsprinzipien und ihrer Materialität letztlich nicht von den Lochkarten aus der Frühzeit der Informations- und Computertechnologie, in denen die jeweiligen Informationen eingestanzt waren. Jedes Speichermedium ist dabei durch Verfall und Verlust gefährdet – wenngleich in verschiedener Weise und in verschiedenem Tempo – indem es physisch degradiert oder entsprechende Hardware zur Auslese nicht mehr zur Verfügung steht. Ein anschauliches Beispiel dafür sind die geradezu heroischen Bemühungen der Beschäftigten der Bundesbeauftragten für die Stasiunterlagen, die auf Magnetbändern und Magnetplatten gespeicherten digitalen Hinterlassenschaften der Stasi und anderer DDR Behörden zu erschließen.<sup>17</sup> Selbst wenn es gelingt, einem obsoleten Speichermedium historische Datensätze zu entreißen, ist damit keineswegs sichergestellt, dass diese dann sinnvoll ausgewertet oder auch nur gelesen werden

---

ries: Mathematics, Alienation, and the Metaphysics of Cyberspace', in: Ders. (Hg.), *Virtual Realities and their Discontents*, Baltimore (ML), London: Johns Hopkins University Press 1996, S. 55-78.

<sup>16</sup> Einzelne Bits auf elektromagnetischen Speichermedien wie Festplatten lassen sich sogar mit einem Elektronenrastermikroskop erkennen und darstellen, vgl. Kirschenbaum, *Mechanisms*, S. 62.

<sup>17</sup> Birgit Salamon, Die archivfachliche Arbeit an den MfS-Geheimdienstunterlagen - Fragen und Herausforderungen, in *Der Archivar*, Jg. 55, 2002, H. 3, S. 203, URL: <https://www.bstu.de/archiv/fachbeitraege/die-archive-der-bundesbeauftragten-fuer-die-stasiunterlagen-bstu-die-archivfachliche-arbeit-an-den-mfs-geheimdienstunterlagen-fragen-und-herausforderungen/>.

können. Denn die oben schon thematisierte Obsoleszenz betrifft nicht nur Hardware, sondern auch Software. Dieses Problem betrifft keineswegs nur digi-born Quellen und entsprechende Anwendungen aus der Frühzeit des Computerzeitalters. Schon Datenbankanwendungen, Textprogramme und Webanwendungen aus den frühen 1990er Jahren lassen sich nur unter größten Mühen lesbar machen, geschweige denn ihr historischer Anwendungszusammenhang und ihr Aussehen (und andere haptische Eigenschaften) wiederherstellen. Die Bemühungen des Marbacher Literaturarchivs, den in jeder Hinsicht obskuren Nachlass Friedrich Kittlers zu erschließen, künden von den vielfältigen Schwierigkeiten eines solchen Unterfangens im Umgang mit born-digitals.<sup>18</sup> Originär digitale Quellen haben somit die paradoxe Eigenschaft, dass sie zwar einerseits abstrakte und digitale Repräsentationen von Informationen darstellen, gleichzeitig aber ihre Materialität und ihre Gebundenheit an spezifische Software die archivalischen Bewahrung und geschichtswissenschaftlichen Analyse maßgeblich beeinflussen. Das jeweilige physische Substrat eines digitalen Objekts seine jeweilige funktional-informationelle Umgebung sind jedoch zugleich die Schlüssel zu seiner Kritik und Analyse als geschichtswissenschaftliche Quelle.

## II. Historische Praxeologie und "material turn": Konzeptionelle und methodische Zugänge

Originär digitale Quellen zu verwenden zwingt zu einer Auseinandersetzung mit ihrem Entstehungskontext, also ihrer Erhebung, der mit ihnen möglichen und beabsichtigten Informationsverarbeitung und Prozessierung und ihrer Rolle und Wirkmächtigkeit in Debatten und Diskursen. Das paradigmatische Beispiel für den Kontext von digi-born Quellen und auch ihr häufigstes informationelles Ökosystem sind Datenbank-Systeme. In der Aneinanderreihung und Sortierung in Klassen und Kategorien, Spalten und Zeilen wurden und werden die meisten Daten aggregiert, gespeichert und verarbeitet. In der Auswahl und Benennung von Attributen und Kategorien einer jeweiligen Datenbank sind wichtige konzeptionelle Vorannahmen und Entscheidungen enthalten, die den einzelnen

---

<sup>18</sup> Der späte Kittler war Linux-Enthusiast und begeisterter Autodidakt im Umgang mit Computern, weshalb seine Verzeichnis und Ablage-logik einzigartig ist und er auch mit selbst geschriebenen Programmen arbeitete. Neben diesen qualitativen Herausforderungen sorgt auch der schiere Umfang des Nachlasses mit gut einem Terrabyte für anhaltende Beschäftigung in Marbach: <https://www.mww-forschung.de/born-digitals>.



Datenpunkten einen Sinnzusammenhang verleihen. Diese Entscheidungen für einen gewisse Repräsentation der Welt und gegen andere, können ihrerseits wiederum Rückwirkungen auf die (reale) Welt haben, indem sie bestimmte Wahrnehmungen und Interpretationszusammenhänge gegenüber anderen begünstigen. Dies sinnstiftenden Wirkung von Datenbanken als eine der zentralen Infrastrukturen der digitalen Informationsverarbeitung hat Rob Kitchin apodiktisch formuliert:

Databases and data infrastructures are not simply neutral, technical means of assembling and sharing data; they are not merely products that store captured data about the world, but are bundles of contingent and relational processes that do work in the world. They are complex socio-technical systems that are embedded within a larger institutional landscape of researchers, institutions and corporations, constituting essential tools in the production of knowledge, governance and capital. Databases are designed and built to hold certain kinds of data and enable certain kinds of analysis, and how they are structured has profound consequences as to what queries and analysis can be performed; their formulation conditions the work that can be done on and through them.<sup>19</sup>

Für die Geschichtswissenschaft und insbesondere die an digi-born Quellen interessierte Digital History ergibt sich daraus eine zwingende quellenkritische Notwendigkeit, sich eingehender mit Datenbanksystemen zu befassen. Laufende und abgeschlossene Forschungsprojekte zu einzelnen Datenbanken und der Kultur-, Ideen- und Wirkungsgeschichte der damit verknüpften Praktiken – des "Databasing" – verweisen gleichzeitig auf das enorme Potential solcher Ansätze.<sup>20</sup> Das Paradox der Virtualität und gleichzeitigen Materialität digitaler Objekte hat der Forschung dabei zusätzlich neue Perspektiven und Zugänge ermöglicht. Der "material turn" und die aus ihm hervorgegangene Medienarchäologie haben einige theoretisch-konzeptionelle Überlegungen zur Grundlage, die wiederum ihrerseits dem Nachdenken über Datenverarbeitung und Datenverwaltung unter digitalen Vorzeichen wichtige Impulse verliehen haben.

---

<sup>19</sup>Rob Kitchin, *The data revolution: big data, open data, data infrastructures & their consequences* (Los Angeles, California: SAGE Publications, 2014), S. 51

<sup>20</sup>Zu praxeologischen Ansätzen in der Zeitgeschichte allgemein siehe: Lucas Haasis und Constantin Rieske, Hrsg., *Historische Praxeologie: Dimensionen vergangenen Handelns* (Paderborn: Ferdinand Schöningh, 2015).

Der ehemalige Leiter der Abteilung für elektronische Dokumente am US-Amerikanischen Nationalarchiv, Kenneth Thibodeau, hat der Archivwissenschaft und der Geschichtswissenschaft eine konzise analytische Definition digitaler Objekte zur Verfügung gestellt. Nach dieser vielzitierten Definition hat jedes digitale Objekt spezifische physikalische, logische und konzeptuelle Eigenschaften:

A digital object is an information object, of any type of information or any format, that is expressed in digital form. That definition may appear too generic to be of any use in addressing the challenge of digital preservation. But if we examine what it means for information to be expressed in digital form, we quickly come to recognize a basic characteristic of digital objects that has important consequences for their preservation. All digital objects are entities with multiple inheritance; that is, the properties of any digital object are inherited from three classes. Every digital object is a physical object, a logical object, and a conceptual object, and its properties at each of those levels can be significantly different. A *physical* object is simply an inscription of signs on some physical medium. A *logical* object is an object that is recognized and processed by software. The *conceptual* object is the object as it is recognized and understood by a person, or in some cases recognized and processed by a computer application capable of executing business transactions.<sup>21</sup>

In eine ganz ähnliche Richtung entwickelte Matthew Kirschenbaum seine Überlegungen zur Materialität digitaler Artefakte. Er unterscheidet zwischen einer formalen Materialität und einer forensischen. Die formale Materialität entspricht dabei der Klasse der logischen Objekte in der Definition von Thibodeau, also den (temporären) Formen und Eigenschaften, die digitale Objekte im Hinblick und im Austausch mit der jeweiligen Software und dem sie umgebenden informationellen System haben müssen. Die forensische Materialität entspricht dem, was Thibodeau die physikalische Ebene nennt, also die manifeste Markierung auf einer physischen Substanz.<sup>22</sup>

---

<sup>21</sup> Kenneth Thibodeau, 'Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years', in: *CLIR and the Library of Congress, The State of Digital preservation: An International Perspective*, April 2002, URL: <https://www.clir.org/pubs/reports/pub107/thibodeau/>.

<sup>22</sup> Kirschenbaum, *Mechanisms*, 13-17.

Von großer theoretischer Bedeutung ist auch die von Niels Brügger eingeführte analytische Differenzierung zwischen (retro-) digitalisierten Quellen, born-digital und re-born-digital Quellen. Vor allem die letzte Gattung hat weitreichende analytische Konsequenzen, da sie auf den Umstand verweist, dass digi-born Quellen, welche zu Forschungszwecken aus ihrem Entstehungskontext extrahiert und zu Analyse aufbereitet und angereichert werden, letztlich mit den Originalen nicht mehr identisch sind. Das beste Beispiel und zugleich Brüggers Gegenstand ist das archivierte Web; aber auch in moderne Formate migrierte historische Dateien entsprechen dieser Definition.<sup>23</sup>

Zu einer Fundierung der tendenziell zum Assoziativen neigenden Medienwissenschaften hat schliesslich Paul Dourishs Untersuchung der Auswirkungen von Materialität(en) auf den Inhalt von Daten und ihre Potentiale zur Weiterverarbeitung (die konzeptuelle Ebene, in Thibodeaus Terminologie) beigetragen. In Fallstudien zu Tabellenkalkulations-Programmen, Datenbanken und der Infrastruktur des Internets zeigt Dourish, dass sich die logische bzw. formale Materialität von Daten und ihre jeweiligen Eigenschaften sowohl auf der Ebene des physischen/forensischen als auch auf der Ebene des konzeptionellen/semantischen niederschlagen, indem sie durch jeweils spezifische Anforderungen und Möglichkeiten informationelle und auch soziale Praktiken determinieren.<sup>24</sup>

Durchaus in Austausch und Anlehnung an die an konkreten und materiellen digitalen Objekten interessierten Medianarchäologie orientiert sich eine jüngere Strömung der Computergeschichte hin zu eher sozialgeschichtlich inspirierten Studien zur Interaktion humaner historischer Akteure mit den jeweiligen Systemen und Maschinen und damit weg von Meistererzählungen genialer (meist männlicher) Erfinder und Pioniere und den von ihnen geschaffenen Technologien.<sup>25</sup> Nicht zufällig geraten durch diese Perspektivverschiebung vermehrt Frauen und ihr Beitrag zur Entwicklung moderne Informations- und Kommunikationstechnologien in den Blick.<sup>26</sup> Zudem entfaltet der Blick weg vom Computer hin zum ihn bedienenden und wartenden Personal (und seit der

---

<sup>23</sup> Niels Brügger, *The Archived Web. Doing History in the Digital Age*, Cambridge, MA: MIT Press, 2018, S. 21-23.

<sup>24</sup> Dourish, *The Stuff of Bits*, 202-206.

<sup>25</sup> Programmatisch formuliert bei: Thomas Haigh, 'Introducing the Early Digital', in *Exploring the Early Digital*, hg. von Thomas Haigh (Cham: Springer International Publishing, 2019), S. 1-18, doi.org/10.1007/978-3-030-02152-8\_1.

<sup>26</sup> Marie Hicks, *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*, History of Computing (Cambridge, MA London, UK: MIT Press, 2018); Janet Abbate, *Recoding Gender: Women's Changing Participation in Computing* (Cambridge, Mass.: MIT Press, 2017).

Durchsetzung des Personal Computers auch privaten Anwenderinnen und Anwendern) ein Panorama der Arbeitswelt und Kultur im Informationszeitalter, das weit über die Labore und Büros des Silicon Valley mit ihren obligaten Hängeecken und Tischtennisplatten hinausreicht.<sup>27</sup>

Um die Funktionsweise historischer Informationsverarbeitungssysteme und die Interaktion daran beteiligter Personengruppen besser verstehen zu können, ist es mitunter hilfreich, historische Systeme zu reaktivieren und historische Praktiken durch Re-enactment nachzuvollziehen. Diese Vorgehensweise immunisiert gegen teleologische Rückprojektionen, in denen historische Phänomene zu bloßen Vorläufern und Stationen auf dem Weg zur Gegenwart (und damit ihrer Vollendung) degradiert werden.<sup>28</sup> Insbesondere Forschungen zur Frühgeschichte der Digitalisierung profitieren dabei vom Austausch mit technischen Sammlungen und Museen, um ein Verständnis für Computer zu entwickeln, die keine der heute gängigen Schnittstellen wie grafische Benutzeroberflächen und Computer-Maus, ja nicht einmal Bildschirme aufwiesen. Eine direkte Auseinandersetzung mit der historischen Technik ermöglicht auch Einsichten in die für ihr Funktionieren notwendigen Arbeiten und die Grenzen mit solcher Technik möglichen Datenverarbeitung. Selbstverständlich müssen Forschungen in diese Richtung eine breite Quellenbasis mit einbeziehen, wie etwa technische Dokumentationen, Notizen, Fotografien, Filmaufnahmen, aber selbstverständlich auch die Produkte der Datenverarbeitungssysteme. Wo das Re-enactment auf die Interfaces und die haptischen Erfahrungen der Datenverarbeitung abzielt – also die physische Materialität ("look-and-feel"),<sup>29</sup> interessiert sich die Emulation für Performance und Funktion, bzw. Fehler und Bugs historischer Software, also die formal-logische Materialität. Die Simulation historischer Verarbeitungsprozesse und Routinen auf modernen Systemen erlaubt ein vertieftes Verständnis nicht nur für das Zustandekommen von Resultaten, sondern gerade auch für die Begrenztheit und Fehleranfälligkeit von Datenverarbeitung. Die Leistungsfähigkeit historischer Datenbanken lässt sich zum Beispiel besser

---

<sup>27</sup> David N. Pellow und Lisa Sun-Hee Park, *The Silicon Valley of dreams: environmental injustice, immigrant workers, and the high-tech global economy*, Critical America (New York: New York University Press, 2002); Nathan Ensmenger, *The Computer Boys Take over: Computers, Programmers, and the Politics of Technical Expertise*, History of Computing (Cambridge, Mass.: MIT Press, 2010).

<sup>28</sup> Thomas Haigh, Mark Priestly und Chrispin Rope haben mit ihrer Studie zur Funktionsweise (bzw. deren häufiges Ausbleiben) des "ersten modernen Computers" ENIAC vorbildlich demonstriert, dass ein detaillierter Nachvollzug und eine entsprechende "dichte Beschreibung" ein große Zahl weit verbreiteter Mythen aufzudecken und zu dekonstruieren vermag: Thomas Haigh, Mark Priestley, und Crispin Rope, *ENIAC in Action: Making and Remaking the Modern Computer*, Cambridge MA: MIT Press 2018.

<sup>29</sup> Lori Emerson, *Reading writing interfaces: from the digital to the bookbound*, Electronic Mediations 44 (Minneapolis: University of Minnesota Press, 2014).

einordnen, wenn sie in der Sprache und Logik ihrer ursprünglichen Abfrage-Routinen benutzt wird. Das Potential der hier skizzierten Zugänge soll nun im Folgenden anhand einer konkreten Fallstudie illustriert werden.

### III. Fragmente einer Kriegsmaschine

Als der 1961 ins Amt gekommenen US-Präsident John F. Kennedy den Ford-Manager Robert S. McNamara zum neuen Verteidigungsminister ernannte, war dies für viele Beobachter, aber auch für den Berufenen selbst eine Überraschung. Der Überraschungseffekt erfasste auch auf die militärische Führung der USA, denn der neue Minister begann sogleich mit einer rigorosen Umstrukturierung der Streitkräfte, die alsbald als revolutionär bezeichnet wurde.<sup>30</sup> Unter dem Eindruck des rasch eskalierenden Katen Krieges und der thermonuklearen Bedrohung setzte McNamara seine Kompetenz und Macht ein, um dem US-Militär nicht nur strategische Überlegenheit zu sichern, sondern auch Effizienz und durchaus auch finanzielle Rentabilität. Bei der Durchsetzung dieser Ziele gegen den nicht unerheblichen Widerstand der militärischen Elite unterstützte ihn ein Stab aus engen Weggefährten aus der Harvard Business School und dem Ford Management, die in der Behörde bald als "whiz kids" bezeichnet wurden (was sich annäherungsweise mit "Klugscheißer" übersetzen ließe). Ähnlich rigoros wie McNamara zu Beginn seiner Amtszeit Armee und Luftwaffe einer strengen Kosten-Nutzen-Kontrolle unterworfen und strategische Machtprotektion logistisch interpretiert hatte, machte er auch den Vietnamkrieg 1967 zur Chefsache, nachdem das seit 1965 stetig intensivierte Engagement der US-Armee nicht zur erwünschten Stabilisierung des südvietnamesischen Regimes und der Eindämmung der kommunistischen Guerilla geführt hatte. Als typischer Manager seiner Zeit und glühender Verfechter des Operations Research und der Systemanalyse verlangte er ein "objektives" Bild der Verhältnisse, auf das künftige politische und militärische Entscheidungen gegründet sein sollten.<sup>31</sup> Wenig verwunderlich sollte diese Objektivität mit den modernsten Methoden der Datensammlung und Analyse – also mit modernen Computern –

---

<sup>30</sup> Charles R. Shrader, *History of Operations Research in the United States Army, V. 2, 1961-1973* (U.S. Department of Defense, U.S. Army Center of Military History, 2008).

<sup>31</sup> Joy Rohde, *Armed with expertise: the militarization of American social research during the Cold War, American institutions and society* (Ithaca: Cornell University Press, 2013); Sharon Weinberger, *The Imagineers of War: The Untold Story of DARPA, the Pentagon Agency That Changed the World*, 2018; Torben Lütjen, 'Vom "Gospel of Efficiency" zum "War of Ideas". Zum Verhältnis von Wissenschaft, Politik und Ideologien in den Vereinigten Staaten', *Archiv für Sozialgeschichte* 50 (2010), S. 373-394.

erreicht werden. Schon vor 1967 setzte die US-Armee in Vietnam Computer ein, etwa zur Verwaltung ihrer Logistik oder der Abwicklung der Soldzahlungen und Urlaubsansprüche der Truppe. Neu war an den seit 1967 unter der Ägide des Civil Operations and Revolutionary Development Support, (CORDS, einer unter Leitung des Botschafters Robert Kromer stehende zivil-militärischen Super-Behörde zur Koordination der Aufstandsbekämpfung<sup>32</sup>) aufgebauten Datenbank-Systemen, dass sie nicht mehr nur militärische Objekte erfassen sollte, sondern nichts weniger als das politische, moralische und psychologische Klima in der südvietnamesischen Zivilbevölkerung. Dazu wurde ein Index sämtlicher 20.000 Dörfer und Weiler unter der Kontrolle des südvietnamesischen Regimes und seiner Alliierten erstellt, inklusive demographischen Informationen und geografischen Koordinaten, Einschätzungen zur Sicherheitslage, und Indikatoren zur jeweiligen sozio-ökonomischen Situation, Selbstverwaltung, Entwicklungspotentiale und weitere. Zusätzlich zu diesen im sogenannten Hamlet Evaluation System (HES) erfassten Grunddaten erhoben speziell geschulte Teams regelmäßig Daten zu den politischen Einschätzungen, Zukunftserwartungen und Moralvorstellungen der Bevölkerung, anhand derer die Wirksamkeit einzelner Maßnahmen militärischer Repression und sozio-ökonomischer Entwicklungsinitiativen evaluiert und somit gezielter dosierbar gemacht werden sollten.<sup>33</sup> Schon während des Krieges war das System massiver Kritik ausgesetzt, es galt als Beispiel einer in-humanen und technisch-kalten Vorgehensweise – wie sie sich etwa in den "body-count"-Statistiken paradigmatisch zeigte, oder aber auch als ineffizient, weil es etwa die Tet-Offensive nicht hatte vorhersehen können oder insgesamt ein zu positives Bild des Kriegsverlaufs produzierte. Die Geschichtswissenschaft hat diese Wahrnehmungen später zwar etwas differenziert, insgesamt aber bestätigt.<sup>34</sup> Auffällig ist jedoch, dass sich solche Studien zwar auf diverses Akten- und anderes Quellenmaterial stützten, sich aber bislang wenig mit den überlieferten Datensätzen, geschweige denn mit dem gesamten Systems auseinandersetzten.

---

<sup>32</sup> Richard A. Hunt, *Pacification: the American struggle for Vietnam's hearts and minds* (Boulder: Westview Press, 1995); Michael E. Latham, *Modernization as ideology: American social science and „nation building“ in the Kennedy era*, New Cold War history (Chapel Hill: University of North Carolina Press, 2000).

<sup>33</sup> Das Datenbanksystem war weitaus umfassender, als hier dargestellt. Es enthielt auch fortlaufend aktualisierte Datensätze zur Präsenz des Vietcong, Tabellen zu dessen zivilen Kommissaren, der regionalen Performance der südvietnamesischen Streitkräfte, etc.

<sup>34</sup> Gregory A. Daddis, *No sure victory: measuring U.S. Army effectiveness and progress in the Vietnam War* (New York: Oxford University Press, 2011).

Die Überlieferungsgeschichte der im Vietnamkrieg eingesetzten Datenbanksysteme wäre einer eigenen Studie würdig. Eine Kartellrechtliche Klage von vier US-amerikanischen Computerherstellern gegen die scheinbare Bevorzugung des marktdominierenden Giganten IBM bei der Vergabe von Regierungsaufträgen im Jahr 1969 führte dazu, dass ein Gericht die Zerstörung sämtlicher elektronischen Dokumente verbot. Während in den leitenden Bürokratien in Regierung und Militär wenig Verständnis für den wissenschaftlichen Wert und die Geschichtsmächtigkeit aberhunderter Meter nur von Maschinen lesbarer Magnetbänder vorherrschte (und auch die offizielle Politik des Nationalarchivs vorsah, wichtige Dokumente auszudrucken und die Originale zu vernichten), empfahl eine von der RAND-Cooperation durchgeführte Studie deren Bewahrung für künftige historischen Forschungen. Dieser Empfehlung wurde letztlich entsprochen und die weit verstreuten Bestände in das Nationalarchiv überführt. Ein dort ehemals mit diesen Beständen befasster Archivar, Donald Fisher Harrison, berichtete allerdings 1988, dass die Forschenden Zurückhaltung bei der Nutzung dieser Bestände zeigten.<sup>35</sup> Neben dem schlechten Ruf der Daten habe dies vor allem an ihrem Format gelegen: die Daten waren für eine bestimmte Software konfiguriert und formatiert (logische/formale Materialität), ein Datenbank-Management-System (DMBS) namens National Information Processing System (NIPS). Um die Nutzung zu erleichtern und auch die bis dato auf Magnetbändern (physische Materialität) gespeicherten Dateien leichter transportabel und übertragbar zu machen, begann das Nationalarchiv 1978 einige der als am wichtigsten Dateien in Software-unabhängige Formate zu übertragen (re-born digitals). Diese zwei Manifestation digitaler Quellen – einmal born digitals, das andere Mal re-born – unterscheiden sich voneinander in mehr als nur ihrem Format (also ihrer formalen, logischen Materialität). Der Bitstream der ursprünglichen NIPS-Dateien enthielt in den vorderen Sektoren Codes des DBMS, die zur Dateiverwaltung notwendig waren, wie etwa Time stamps, Prüfsummen und andere Kontrollinformationen für die Programmierer. Im Prozess des Transfers in modernere Formate wurden dies Sektionen entfernt und in Dokumentation ausgelagert, jedoch nur teilweise. So lässt sich nach der Migration zwar der ursprüngliche konzeptuell/semantische Inhalt der Dateien erschließen, nicht aber ihr logisch-formaler.

---

<sup>35</sup> Donald Fisher Harrison, 'Computers, Electronic Data, and the Vietnam War', *Archivaria* 26 (1988), S. 18–32.

Das Datenbanksystem des CORDS in Vietnam ist aber noch in weiteren Manifestationen überliefert. Von der Erhebung der Daten in den Dörfern in Weilem Südvietnams bis zu ihrer Übertragung in die Zentralcomputer des Armeekommandos in Saigon und schließlich in jene des Pentagons durchliefen die Informationen mehrere Stationen und wurden dabei oftmals auf neue physische Speichermedien übertragen, die wiederum jeweils spezifische materielle Eigenschaften und spezifische formale Anforderungen hatten. Am Beispiel des HES lässt sich illustrieren, wie durch den Nachvollzug des Lebenszyklus der Daten und ihre mehrfachen konzeptuellen, logischen und physischen Transformationen die jeweils beteiligten Akteure hervortreten.

Am Beginn standen auf Karton gedruckte Fragebögen. Die Fragebögen und ihre Auswertungsmatrix waren von einem Team aus Sozialwissenschaftlern um Ithiel de Sola Pool entwickelt worden, der durch die Computerbasierte Berechnung von Wahlergebnissen Bekanntheit erlangt hatte.<sup>36</sup> Die Karten bestanden aus insgesamt 136 Fragen zu 18 Indikatoren aus den Bereichen Politik, Sicherheit, Wirtschaft, Verwaltung und Gesundheit, die jeweils in einer fünfstufigen Skala zu bewerten waren. Teams aus jeweils einem amerikanischen Offizier (District Senior Advisor, DSA) und einigen Vietnamesischen Assistenten und Übersetzern wurden in jedem der 256 militärischen Sub-Distrikten monatlich ausgeschiedt, um die Situation in allen Dörfern und Weilem zu erfassen und auf den Karten einzutragen. Am Ende des Monats musste der DSA die Ergebnisse seiner Erhebungen inklusive zusätzlicher Bemerkungen auf einem vorgedruckten Formblatt eintragen und an das Provinzhauptquartier weiterleiten, wo die Formblätter gesammelt, auf Vollständigkeit und Kohärenz überprüft und schließlich nach Saigon in das CORDS-Hauptquartier weiter geschickt wurden. Dort wurden die Blätter von der Reports and Analysis Division (RAD) erneut geprüft und sortiert. In der Data Management Agency des Oberkommandos wurden die Formblätter schließlich von angestellten Vietnamesinnen auf Lochkarten gestanzt und dann in die Computer eingespeist.<sup>37</sup> Auf Grundlage dieser monatlich aktualisierten Daten

---

<sup>36</sup> Jill Lepore, *IF THEN: How the Simulmatics Corporation Invented the Future* (New York: Liveright, 2020); zur Beteiligung der Sozialwissenschaft am Vietnamkrieg allgemein siehe: Mark Solovey und Hamilton Cravens, *Cold War Social Science: Knowledge Production, Liberal Democracy, and Human Nature* (Basingstoke (GB): Palgrave Macmillan, 2014); Stuart W. Leslie, *The Cold War and American Science: The Military-Industrial-Academic Complex at MIT and Stanford* (New York: Columbia Univ. Press, 1993), sowie Anm. 31.

<sup>37</sup> NARA, MACV/CORDS, Pacification Study Group, HES Briefing manuscript 1972, ARC ID 4489248, Box 31.



berechneten Analysten des RAD (zum Teil Militärs und Geheimdienstler, aber auch zivile Sozialwissenschaftler und Angestellte privater Datenverarbeitungs-Firmen unter Regierungsvertrag) Veränderungen und Prognosen zur politisch-militärischen Situation. Die Resultat wurden schließlich in den monatlich erscheinenden "Southeast Asia Analysis Reports" zusammengefasst, auf riesigen Karten geplottet und der gesamte Monatsreport zusätzlich in mehrfacher Ausführung auf Endlospapier ausgedruckt. Die schiere Menge der im HES verarbeiteten Daten war für damalige Verhältnisse beispiellos: 9.000 Lochkarten wurden monatlich gestanzt, 10 Millionen Einträge enthielt das Masterfile auf mehreren Magnetbändern und 90.000 Seiten umfassten die Reports. Insgesamt produzierte das Oberkommando der US-Streitkräfte in Vietnam sagenhafte 6 Tonnen bedrucktes Papier, weshalb der Vietnamkrieg bis heute als einer der am besten dokumentierte militärische Konflikt des 20. Jahrhunderts gilt.<sup>38</sup> Dieses gewaltige und komplizierte Unterfangen war an verschiedenen Stationen des Prozederes anfällig für Fehler und Verluste, die nicht nur an der mitunter eingeschränkten Zuverlässigkeit der Befragten und an der Erhebung und Aggregation beteiligten Personen, oder an der allgemeinen Unübersichtlichkeit des Guerilla-Krieges lagen, sondern auch an der jeweiligen Materialität der Informationen. Die Frage-Karten für die einzelnen Dörfer schafften es oftmals nicht bis ins Provinzhauptquartier, etwa wenn sie bei Gefechten verloren gingen, durch das Dschungelklima, Regen oder Matsch zerstört wurden oder durch Nagetierfraß beschädigt. Das System erwies sich insgesamt auch oft als zu statisch um die Dynamik des Krieges und seiner Konsequenzen zu erfassen. So wussten sich die Analysten nicht anders zu behelfen, als die Millionen von Binnenflüchtlingen unter der Kategorie ihrer Herkunftsdörfer zu erfassen, was zu das Gesamtbild oftmals bis ins Absurde verzerrte.<sup>39</sup> Entsprechende Revisionen des Systems in den Jahren 1970 und 1971 führten zu weiteren Problemen. So sollte 1970 die Genauigkeit des Systems verbessert werden, indem weitere Fragen auf die Dorf-Karten aufgenommen wurden. Diese erweiterte Matrix passte dann allerdings nicht mehr auf eine einzige Lochkarte, was zu einer Verdoppelung der benötigten Menge an Lochkarten und damit weiteren potentiellen Fehlern und Ungenauigkeiten führte. Eine Revision der Erhebungsmatrix und damit des Umfangs und der Form der Datensätze zog zudem stets

---

<sup>38</sup> Erwin R. Brigham, 'Pacification Measurement in Vietnam: The Hamlet Evaluation System', Military Review, May 1970.

<sup>39</sup> NARA, MACV/CORDS, Reports and Analysis Directorate-Analysis Division, Analysis Report Concepts of HES Reaggregation for 1971, 14 December 1970, ARC ID 6087884, Box4, folder 472-10112-4-3.

eine aufwändige Neuprogrammierung des Systems nach sich, da Abfrage-Codes und Routinen der Verarbeitung neu geschrieben werden mussten. Die Computer selbst schließlich, Raumfüllende IBM 360 Maschinen, waren enorm wartungsintensiv und fielen durch die Feuchtigkeit und die häufigen Stromausfälle oft tagelang aus. Die Bandspulen, die das Magnetband mit dem Masterfile bewegten brannten durch und nicht alle Magnetbänder waren der hohen Belastung und dem tropischen Klima gleich gut gewachsen.

#### Fazit

Was lässt sich mit einem historischen Datenbanksystem und seinen heterogenen Spuren machen und was nicht? Die migrierten HES-Dateien im US-Nationalarchiv sind relativ vollständig und enthalten Informationen zur demografischen und sozio-ökonomischen Situation in den ländlichen Gebieten Südvietnams während des Krieges, wie sie sonst in dieser Fülle schwer zu finden sind. Vor allem ist den einzelnen Datensätzen stets eine geografische Information (UTM) vorangestellt. Es ist also sehr verlockend, sich diesen Quellen mit dem Standard-Bestock der Digital History anzunähern und sie auf Karten zu projizieren und Dynamiken und Veränderungen in Histogrammen und Zeitreihen zu visualisieren. Abgesehen davon, dass dies ein sehr unkritisch-positivistischer Zugang wäre, würde er ironischerweise exakt jene Schritte und Darstellungsformen replizieren, welche die Akteure selbst angewendet hatten, um ihre Pazifizierungspolitik zu evaluieren, aber auch um repressive Maßnahmen wie Bombardements zu legitimieren. Ein positivistischer Zugang verbietet sich somit aus analytischen und ethischen Gründen. Zudem offenbart eine gründliche quellenkritische Auseinandersetzung mit dem Datenbanksystem, seinen Revisionen und der jeweiligen Determiniertheit der Daten durch die konzeptuelle, logische und physische Materialität die Unzulänglichkeit und den konstruierten Charakter des Systems. Dies hat renommierte Sozialwissenschaftler nicht davon abgehalten, die HES-Daten in das empirische Fundament weitreichender Theorien einzuweben.<sup>40</sup>

Dagegen scheint eine Annäherung an die Frage, ob das US-Militär ausreichend über die Situation und Stimmung der Landbevölkerung informiert war oder nicht, analytisch fruchtbarer. Ob also eine ausreichend akkurate Repräsentation

---

<sup>40</sup> Stathis N. Kalyvas und Matthew Adam Kocher, „The Dynamics of Violence in Vietnam: An Analysis of the Hamlet Evaluation System (HES)“, *Journal of Peace Research* 46, Nr. 3 (2009), S. 335-355, <https://doi.org/10.1177/0022343309102656>.

des politisch-militärischen Geschehens durch das System produziert wurde, wie einer der maßgeblichen Analysten in seinen späteren Schriften insistiert<sup>41</sup>, oder ob die politisch und militärische Führung auf eine Projektion hereinfließen, die sie selbst konstruiert hatten, ist eine historiografisch und politisch relevante Frage. Angesichts der enormen Bedeutung, welche die Erfahrungen mit dem HES-System für die US-amerikanische empirische Sozialforschung und die Weiterentwicklung von Datenbanktechnologien hatten, erscheint die Auseinandersetzung mit der Praktik des "Databasing" und seinen Ideen- und Wissensgeschichtlichen Auswirkungen ebenfalls als ein produktiver Zugang.

Die Analyse der HES-Datenbank und der damit verbundenen Systeme verspricht durch seine spezifische Überlieferungsgeschichte auch weiterreichende Erkenntnisse für die archivalische Bewahrung und geschichtswissenschaftlicher Erforschung historischer Datenbanken: Auf der Ebene der Archive spricht die Erfahrung mit diesem System dafür, eine möglichst umfassende Bewahrungspraxis anzustreben, also nicht nur historische Datensätze der Forschung zugänglich zu machen, sondern wenn möglich die vollständige Umgebung etwa in Form von emulierten Systemen inklusive extensiver Dokumentation zu bewahren.<sup>42</sup> Die Geschichtswissenschaft und hier wiederum insbesondere die Digital History wäre gut beraten, sich in engen Austausch mit Zugängen wie der experimentellen Medianarchäologie zu begeben, um ihre eigentliche Kernkompetenz, die Kritik und Kontextualisierung von historischen Quellen, auch im digitalen Zeitalter nutzbar zu machen.<sup>43</sup>

---

<sup>41</sup> Thomas C Thayer, *War Without Fronts*. (Naval Institute Press, 2016).

<sup>42</sup> Matthew Kirschenbaum et al., 'Digital Materiality: Preserving Access to Computers as Complete Environments', *University of California Digital Library* (2009), <https://escholarship.org/uc/item/7d3465vg>; Loebel, Jens-Martin, *Lost in translation: Leistungsfähigkeit, Einsatz und Grenzen von Emulatoren bei der Langzeitbewahrung digitaler multimedialer Objekte am Beispiel von Computerspielen* (Glückstadt: Hülsbusch, 2014); Dirk von Suchodoletz, *Funktionale Langzeitarchivierung digitaler Objekte - Erfolgsbedingungen des Einsatzes von Emulationsstrategien*, nestor edition 1, Freiburg im Breisgau: Albert-Ludwigs-Universität, 2008, <http://nbn-resolving.de/urn:nbn:de:0008-2008070219>.

<sup>43</sup> Gerben Zaagsma, 'On Digital History', *BMGN - Low Countries Historical Review* 128, Nr. 4 (2013), URL: <https://doi.org/10.18352/bmgn-lchr.9344>; Fickers, Andreas, 'Hands-on! Plädoyer für eine experimentelle Medienarchäologie', *Technikgeschichte* 82,1 (2015), S. 67-85.