

Annotation, Simulation und Analyse eines historischen Datenbanksystems

Moritz Feichtinger¹

Abstract: Seit den 1960er Jahren verwenden Regierungen und Behörden der westlichen Welt digitale Kommunikations- und Informationstechnologien. Damit die Geschichtswissenschaft und damit auch die Bürgerinnen und Bürger ihrer demokratischen Rolle der Aufarbeitung und Kontrolle nachkommen können, ist es wichtig, Standards zur Bewahrung und Analyse digitaler Artefakte des Regierungshandelns zu entwickeln. Dieses Paper zeigt die Schwierigkeiten eines solchen Unterfangens anhand der Überlieferung von Datenbanksystemen aus dem Vietnamkrieg exemplarisch auf. Seit 1966 setzte die US Armee in Vietnam Systeme zur automatisierten Sammlung und Auswertung von Daten für ihre strategische und politische Planung ein. Zwar sind einzelne dieser Datensätze überliefert, jedoch teilweise in einer obsoleten Formatierung, teilweise mit nur fragmentarischer Dokumentation. Der eigentliche Quellcode des Database Management Systems, der Rückschlüsse auf die Funktionsweise des historischen Systems erlauben würde, ist nicht überliefert. Das Paper erläutert diese Situation und diskutiert die Möglichkeit, aus den überlieferten Datensätzen und der Dokumentation eine kritische Simulation der Funktionsweise des historischen Systems zu entwickeln.

Keywords: Geschichtswissenschaft; Informationswissenschaft; Archivwissenschaft; Computergeschichte; Medienarchäologie; Forensik

1 Die Herausforderung durch *digital born* Objekte

Die Geschichte der zweiten Hälfte des 20. Jahrhunderts und des Beginns des 21. Jahrhunderts läßt sich ohne Einbezug digitaler Quellen und Artefakte ("born digitals"), insbesondere digitaler Datenbanken, nicht mehr verstehen und beschreiben. Datenbanken sind zentrale Infrastrukturen digitaler Informations- und Kommunikationstechnologien (ICT). Ihre Struktur bestimmt, welche Informationen erhoben, gespeichert und verarbeitet werden, und wie sie zueinander in Beziehung gesetzt werden können. Seit den 1960er Jahren bilden die jeweiligen Architekturen und Funktionsweisen von Datenbankmanagement Systemen (DBMS) die Grundlage für die Produktion, Ordnung und Weiterverarbeitung von Wissen in digitalen ICTs.[Ha09]. Datenbanken sind weit mehr als bloße Container von Informationen. Sie sind komplexe soziotechnische Systeme, deren Architektur die Generierung und Repräsentation von Wissen im digitalen Zeitalter maßgeblich beeinflussen.[Ki14] Um die vielfältigen Elemente und Auswirkungen der Entwicklung und Nutzung von Datenbank-Technologie historisch fassen zu können, bietet es sich daher an, über das technische Objekt hinaus zu gehen und epistemologischen Vorbedingungen, die Praktiken der Interaktion mit dieser

¹ Universität Basel, Departement Geschichte, Hirschgässlein 21, 4051 Basel, Schweiz moritz.feichtinger@unibas.ch

Technologie, ihre materiellen sowie informationelle Umgebung und die Rückwirkungen auf die Welt mit in die Analyse einzubeziehen. Das Verb *“databasing”* drückt dieses holistische Verständnis aus, da es den Umgang mit Datenbank-Technologie als Praxis betont.[Do17] *Databasing* kann und sollte als soziale und kulturelle – und damit historisch verortete – Praxis verstanden werden. Diese Perspektive entspricht den neueren Ansätzen der Geschichte von Informationstechnologie, in denen nicht mehr die Technik, “der Computer” als Ideal im Singular, sondern “Computing”, also die vielfältigen Praktiken der Interaktion und Nutzung im Fokus stehen. Diese Perspektive hat es ermöglicht, den Umgang mit Informationstechnik als soziale Praktik zu verstehen und somit die beteiligten Personen und Gruppen genauer zu analysieren und den Fokus der “klassischen” Computergeschichte auf männliche Erfinder und Geschäftsleute in westlichen Ländern zu überwinden. [Hi18], [Mc20], [Ch21] Die etablierten geschichtswissenschaftlichen Zugänge müssen auf ihre Übertragbarkeit auf born digitals überprüft und gegebenenfalls erweitert werden. Die Grundlage der wissenschaftlichen Untersuchung von Zeugnissen der Vergangenheit bildet die Quellenkritik, also die Überprüfung der Echtheit, Vollständigkeit, Herkunft und Glaubwürdigkeit von Überlieferungen und Hinterlassenschaften. Aber sind hergebrachte Konzepte etwa der Originalität und Authentizität sinnvoll anwendbar auf digitale Artefakte angesichts ihrer einfachen Kopierbarkeit?[Fi21] Wie könnte in diesem Zusammenhang der Nachvollzug der Provenienz (also Herkunft) von digitalen Quellen aussehen? Kann im Falle von kollaborativ oder interaktiv entstandenen Quellen noch von Autorschaft gesprochen werden? Um diese neue Quellengattung bearbeiten zu können, müssen Methoden aus der Medianarchäologie und der Computer-Forensik in den Werkzeugkasten der Geschichtswissenschaft aufgenommen werden.[Ri22] Auch das analytische Instrumentarium der Geschichtswissenschaft, das verstehende Erschließen (Hermeneutik) muss an die spezifischen Eigenschaften von born digitals angepasst werden. Die Frage nach dem Zusammenhang zwischen der Medialität und dem (semantischen) Inhalt erhält bei digitalen Objekten ein noch größeres Gewicht als bei nicht-digitalen, da digitale Objekte nicht einfach so gelesen werden können, sondern eine passende Softwareumgebung und Computer für den Zugriff auf sie notwendig sind. Klassische Fragen der historischen Wissenschaften etwa nach der Perspektivität und Repräsentativität, dem Kontext und der kulturellen Gebundenheit von Quellen müssen in der Analyse von digital born Objekten ebenfalls bearbeitet werden, erschließen sich aber nicht durch heuristische Methoden aus den Quellen selbst, sondern setzen das Vorhandensein von Kontextinformationen und Dokumentation voraus.[B117] Aus archivalischer Perspektive stellt sich die Frage der Auswahl des zu bewahrenden digitalen Erbes auf zweifacher Ebene: Zunächst ist die Frage der Selektion zu beantworten, also zu entscheiden, welche Quellen und Artefakte der stetig und exponentiell anwachsenden Menge an digitaler Kommunikation und Information für die Nachwelt potentiell von Interesse sein könnten. Dann ist zu klären, wie genau die ausgewählten Zeugnisse zu bewahren sind. Wie ist der technischen Obsoleszenz zu begegnen, also dem Problem, dass sich historische Formate von heutiger (und zukünftiger) Software nicht mehr auslesen lassen? Was genau gehört zu einem digitalen Artefakt, nur seine Repräsentation in Binärcode und die entsprechende Codierung, oder auch die Umgebung im weitesten Sinne, oder sogar Hinweise auf die historische Nutzung? Wie soll der Zugriff auf digitale Objekte in veralteten Formaten gewährleistet werden? Indem sie

in modernere Formate migriert werden oder indem ihre ursprüngliche Funktionsweise in modernen Betriebssystemen emuliert wird?[Lo14], [Bo21]

Das laufende Forschungsprojekt “Ctrl + All – Computing the Social” untersucht ein Datenverarbeitungssystem aus dem Vietnamkrieg entlang der hier vorgestellten analytischen Linien als historisches Fallbeispiel von databasing. Bevor auf die einzelnen Schritte des Forschungsprojektes, erste Ergebnisse und Schwierigkeiten eingegangen werden kann, muss der Entstehungszusammenhang der Datenverarbeitung und die Besonderheiten der archivierten Überreste dargestellt werden.

2 Fallstudie: Databasing im Vietnamkrieg

Die US Armee setzte während ihres Eingreifens in den Krieg in Vietnam (1965-1975) Computer für verschiedene Zwecke ein, für ihre Logistik, Verwaltung, Sammlung von Geheimdienst Informationen, strategische Planung und Evaluation. Auch für die Counterinsurgency, die integrierte politisch-militärische Bekämpfung der Guerilla, setzten die Amerikaner und ihre Verbündeten Computer und automatisierte Datenverarbeitung ein. Die 1967 gegründete Zentralbehörde zur Koordination der zivilen und militärischen Counterinsurgency Massnahmen, das CORDS (*Civil Operation and Revolutionary Development Support*), verwendete Computer und Datenbanken um ihre Projekte zu koordinieren und zu evaluieren. Zwei davon sollten ein monatliches umfassendes Bild der demographischen, sozioökonomischen Situation der ländlichen Bevölkerung produzieren und ihre politischen Ansichten und Haltungen erheben, das *Hamlet Evaluation System* (HES) und das *Pacification Attitudes Analysis System* (PAAS). Das HES war ein Verzeichnis aller 12.000 Weiler und Dörfer in Südvietnam, das die geografische Lage der Siedlungen, soziodemografische Daten über ihre Bewohner und eine fünfstufige Bewertung enthielt, in die der Grad der Kontrolle der jeweiligen Gemeinschaft durch die Regierung oder die Kommunisten angegeben wurde. Die Daten für das HES wurden von 250 US-amerikanischen Militärberatern in allen 44 Provinzen Südvietnams monatlich erhoben und in einer zentralen Datei gesammelt. Das HES wurde seit seiner Inbetriebnahme im Jahr 1967 bis zur Übergabe an die vietnamesische Regierung im Jahr 1973 dreimal überarbeitet und gilt als die umfassendste Datenbank über die Bevölkerungsverteilung und -dichte im Südvietnam der Kriegszeit. Aufgrund seiner Geoinformationen und Bevölkerungsregister diente es als Grundlage für die Verknüpfung vieler anderer von CORDS betriebenen Datenbanken.[Fe22] Das PAAS kann als qualitatives Korrelat zum eher quantitativen HES betrachtet werden. Es bestand aus behavioristischen Umfragen, die von speziell geschulten vietnamesischen Fragestellern monatlich in bestimmten Regionen durchgeführt wurden. Der Fragebogen umfasste Themen wie Lebensmittelpreise, Hoffnungen und Befürchtungen für die Zukunft sowie allgemeine politische Einstellungen und Ansichten. Obwohl PAAS nur auf einer recht kleinen Stichprobe der Bevölkerung basierte, wurden die Erhebungen unter ethnischen und religiösen Minderheiten sowie unter Frauen und damit einem erstaunlich breiten sozialen Spektrum durchgeführt. Die Ergebnisse wurden kodiert und in einer Datenbank zusammengefasst, die

mit dem HES-System kompatibel war. Beide Systeme, HES und PAAS, produzierten monatliche umfassende Ergebnisberichte, deren wichtigsten Erkenntnisse zudem auf Diagrammen, Landkarten und Broschüren visuell dargestellt wurden. Diese Befunde dienten Politikern und Militärs zur Einschätzung der Lage in Südvietnam, zur taktischen und strategischen Planung, aber auch zur Legitimation solcher Manöver sowie des amerikanischen Engagements in Vietnam insgesamt.

2.1 Provenienz, Formate, Dokumentation

Das US-Nationalarchiv (*National Archive and Records Administration*, NARA) erhielt die HES und PAAS Datenbanken (und einige andere) auf 9-Spur Magnetbändern in den späten 1970ern. Da die Abteilung für elektronische Dokumente (*Electronic Records Archive*, ERA) damals nicht über das benötigte Equipment verfügte, extrahierte das National Institute for Health die Daten von den (schnell degradierenden) Magnetbändern. Die Daten waren zum Teil im damals gängigen, IBM-eigenen, EBCDIC Format abgelegt, zum Teil in einem Format namens NIPS: Das *National Information Processing System Formatted File System* war ein aus dem *1400 Formatted File System* (FFS) hervorgegangenes Datenbankformat, das speziell für das US Militär entwickelt worden war. Es lässt sich der Familie der selbstbeschreibenden Datenformaten zuordnen, das heißt, Struktur und Schema der Datensätze war in den vorderen Segmenten der Datensätze selbst beschrieben.[FS76] Um Forschenden den Zugang zu den Datensätzen des Vietnamkrieges zu erleichtern, begann das ERA in den 1980ern damit, die Daten in das ASCII Format zu übersetzen. Dafür wurde eine Software entwickelt (NIPStran), mit der sich grosse Teile der Konversion automatisieren ließ. Die Datensätze liegen also in mehreren Formaten und Speichermedien vor: einige Reports (die Endprodukte beider Systeme) sind als Ausdrücke auf Endlospapier erhalten; das NARA verfügt über Bitstream-Images der kompletten Datensätze in ihrer originären Formatierung (NIPS) und stellt auf seiner Homepage die ASCII-konvertierten Dateien zum Download zur Verfügung. HES und PAAS wurden von spezialisierten Analyse-Abteilungen in mehreren Regierungsbehörden benutzt und analysiert: von der *Research and Analysis Division* des CORDS in Saigon, vom *Combined Intelligence Data Center* der Armee in Saigon, vom *Systems Analysis Office* im *Department of Defense* in Washington, einigen Sozial- wissenschaftlichen Forschungsinstitutionen, sowie von privaten Computerfirmen, die mit Regierungsverträgen die Datenbanken betreuten und weiterentwickelten. Alle diese Institutionen hatten Nutzerhandbücher und haben ihren Umgang mit den Systemen mehr oder weniger umfassend dokumentiert. Das NARA bewahrt zudem einen um fangreichen Bestand der Research and Analysis Division Division des CORDS auf, aus dem sich die Geschichte der Entstehung der Systeme, die Praxis der Datensammlung, sowie mehrere Revisionen der Systeme und die jeweiligen Prozessierungsverfahren zumindest teilweise rekonstruieren lassen. Weitere Quellen und Dokumente im NARA und anderen Archiven lassen Rückschlüsse auf die Verwendung der Daten in der strategischen und taktischen Einsatzplanung sowie der politischen Entscheidungsfindung und Öffentlichkeitskommunikation entscheidender Regierungsbehörden zu.

3 Forschungsdesign

Das Forschungsprojekt zielt darauf ab, ein in Fragmenten überliefertes Datenverarbeitungssystem durch eine Untersuchung der entsprechenden Dokumentationen möglichst breit zu kontextualisieren, mit (quellenkritischen) Metadaten zu annotieren und durch eine (experimentelle) Rekonstruktion und Simulation zu analysieren. Das von den Archivarinnen und Archivaren des ERA im amerikanischen Nationalarchiv entwickelte Programm NIPStran produziert ASCII-formatierte Datensätze als .csv Files und HTML Files, in denen Tabellen-, Spalten- und Attributnamen abgelegt sind. Grundlage für diese Umwandlung sind die *fixed set* der NIPS Dateien, in denen nicht-repetitive Einträge gespeichert sind (also die Kategorien) und jeweils mehrere zugehörige *periodic sets*, also dynamisch Datensätze, in denen die zugehörigen Werte abgelegt waren und monatlich oder vierteljährlich erneuert wurden. Mit diesen Daten lassen sich sehr einfach Tabellen und Data-Arrays erstellen und auswerten. Allerdings enthielten die originalen NIPS-Daten (und auch die in EBCDIC gespeicherten Daten) darüber hinaus in den vorderen Segmenten weitere Informationen, namentlich ein *control set*, das einen Unique Identifier enthielt und zur Verknüpfung der Datensätze genutzt werden konnte, ein *variable set* für Kommentare, File-Maintenance Statements und Log-files, sowie weitere Strukturdaten, die der Interaktionen mit verschiedenen Prozessoren und Verarbeitungsverfahren dienten.

3.1 Extraktion von Strukturinformationen

Der erste Schritt der Analyse besteht darin, ein Verfahren zu entwickeln, das technisch-funktionale Metadaten aus den ersten Segmenten des Bitstreams extrahiert. Grundlage dafür ist der Source-Code von NIPStran und frei verfügbare Programme zur Migration von EBCDIC Dateien. Dabei geht es darum zu klären, ob sich aus dem archivierten Bitstream Aussagen über Funktionsweise und Architektur des Systems ziehen lassen. Bestenfalls lassen sich Verfahren entwickeln, mit denen sich Strukturinformationen und Dateischemata nicht nur aus verschiedenen Datensätzen im NIPS-Format auslesen, sondern das Verfahren so konzipiert werden, dass sich auch aus Daten in verwandten Formaten der FFS Familie Schemata extrahieren lassen. Das angestrebte Ziel dieses Schrittes besteht im Entwurf eines RDF Knowledge-graphs, in dem die Datensätze in ihrer Gesamtheit gemeinsam mit den technischen Strukturinformationen abgelegt sind. Damit soll in einem späteren Schritt die Funktionsweise der Systeme nachvollzogen (und kritisiert) werden können.

3.2 Annotation mit Metadaten

Die im ersten Schritt gewonnenen Strukturinformationen müssen nach einem einheitlichen Schema abgelegt werden und als Metadaten zu den (konvertierten) Daten in den RDF Graph hinzugefügt werden. Zur Beschreibung und Ablage der technisch-strukturellen Informationen wird der Metadatenstandard PREMIS verwendet. Das Vokabular von PREMIS

(PREservation Metadata: Implementation Strategies), dessen Data Dictionary in der letzten Version 3.0 2016 publiziert wurde, hat sich in den letzten Jahren als de facto Standard für Metadaten zur Beschreibung digitaler Objekte etabliert. Mit PREMIS können die strukturellen und technischen Eigenschaften von historischen Datenbanksystemen erfasst werden. Zudem kann PREMIS, das vier große Klassen von Einheiten (entities) kennt - *Object*, *Event*, *Agent*, *Rights* - die Interaktion sowohl von Personen als auch von technischen Systemen mit Datenbanken abbilden. Die Einheit *Object* hat drei Unterkategorien *file*, *bitstream* und *representation*, die geeignet sind, den Prozess der Datenverarbeitung und die jeweiligen Soft- und Hardware Umgebungen zu beschreiben. Die Einheitsklassen *Event* und *Agent* können jegliche Interaktion mit dem System ausweisen, da *Agent* sowohl Maschinen oder Software als auch Personen bezeichnen kann. Die Daten des HES und PAAS durchliefen jeweils einen mehrstufigen Lebenszyklus, in dessen Verlauf sich ihr Inhalt, ihr Ablage-Format, die entsprechenden Speichermedien und ihr Verarbeitungs- und Verfeinerungsgrad veränderte. Zudem konnten sich die Datensätze in untereinander in den vorgenannten Aspekten unterscheiden. In den Archivwissenschaften hat sich die Erstellung von Data-Lifecycle-Modellen zur Beschreibung von Daten etabliert. Um den Lebenszyklus der Datensätze zu ermöglichen, muss ein formalisiertes Schema entwickelt werden, in dem sich die jeweilige Provenienz-Information zu den Datensätzen für die jeweiligen Stufen in ihrem Lebenszyklus beschreiben lassen. Damit soll zum Beispiel festgehalten werden, wo, von wem, mit welcher Methode und auf Grundlage welchen konzeptionellen Modells Daten erhoben wurden. Auch soll sich damit festhalten lassen, welchen Grad der Weiterverarbeitung, Säuberung, Aggregation oder Verknüpfung mit anderen Daten die jeweiligen Datensätze aufweisen.

Die Data Documentation Initiative (DDI) hat mit dem Discovery Vocabulary einen Metadaten-Standard entwickelt, mit dem sich sozialwissenschaftliche Studien im Hinblick auf die vorgenannten Schritte in ihrem Lebenszyklus beschreiben lassen. Das Ziel des zweiten Schrittes besteht darin, eine quellen-kritische Beschreibung für die Daten aus dem Vietnamkrieg aus dem Vokabular von DDI-Discovery zu entwickeln und in den PREMIS-basierten Knowledge-graph aus Schritt 1 zu integrieren. Damit wird dieser nicht nur um die Dimension einer (kritischen) Feststellung der Provenienz Informationen erweitert, sondern es lassen sich erste Korrelationen zwischen technischen und Provenienz-Metadaten erheben. So lassen sich, wenn sämtliche überlieferte Daten einheitlich annotiert und erfasst sind, die Kohärenz zwischen verschiedenen Versionen überprüfen und die Auswirkungen verschiedener Aggregationsstufen feststellen. Die Anwendung von Metadaten-Standards wie DDI-Discovery und PREMIS wird auch deren Grenzen und entsprechenden Bedarf an deren Ausbau für historische Datenbanken und Software aufzeigen. Beide Standards kommen aus der Archivwissenschaft und dienen vornehmlich der Dokumentation, dem Austausch und der Zugänglichmachung von digitalen Objekten, nicht deren Analyse. Für eine kritische Erforschung nach den Standards der Geschichtswissenschaft ist gerade die Ausweisung von Lücken und Unklarheiten in Provenienz und Integrität von Bedeutung und ein entsprechende Auszeichnungs-Vokabular muss entwickelt und erprobt werden. Bestenfalls kann das Forschungsprojekt dafür Anhaltspunkte und Erfahrungen generieren.

3.3 Simulation der Funktionsweise

Der finale und wichtigste Schritt des Forschungsprojekts besteht aus einer kritischen Simulation der Funktionsweise der historischen Systeme. Damit soll überprüft werden, ob sich das experimentelle Reenactment, das in den Science and Technology Studies ein erprobtes Verfahren ist, in den Methodenkanon der reflexiven Computergeschichte und Digital History integrieren lässt.[Du20], [He20], [FPS16]

Die Simulation umfasst auf der Ebene des Gesamtsystems eine formale Modellierung der Datenverarbeitungsprozesse, inklusive der zugrundeliegenden Problemstellungen, der beteiligten Personen, Maschinen, Medien und Materialien sowie der einzelnen Systemkomponenten und ihr Ineinandergreifen auf Softwareebene. In Kombination mit den technischen Metadaten soll somit die Funktionsweise des Gesamtsystems nachvollziehbar und Wechselwirkungen zwischen Materialien, Formaten, und Inhalten überprüfbar gemacht werden. Die zugrundeliegende Idee ist, einen Graph zu erstellen, der den gesamten Lebenszyklus der Daten im System zusammen mit der Funktionsweise und Interaktion mit den einzelnen Komponenten repräsentieren kann. Damit soll auch ermöglicht werden, die jeweiligen Effekte der Veränderung einzelner Parameter und Komponenten auf die Performance des Systems und die Validität der durch das System generierten Informationen zu testen. So lässt sich nachvollziehen, ob sich statistische und konzeptuelle Verzerrungen und Vorannahmen im Prozess der Datenverarbeitung verschärften, oder ob sie durch Aggregationsverfahren nivelliert wurden. Nicht zuletzt sollte die kritische Simulation überprüfbar machen, wie sich die Ergebnisse des Systems zu Lageberichten und Einschätzungen zum Kriegsverlauf aus anderen Quellen verhielten. Das Ziel der Simulation auf Softwareebene besteht darin, die Abfrage- und Systemroutinen in moderner Software nachzubauen. Einerseits sollen verschiedene Query-Prozesse in SPARQL nachgebaut werden, um die historischen Abfrage-Routinen nachvollziehbar zu machen, andererseits soll eine Grundlage für die Entwicklung weiterer Abfrage-Kombinationen ermöglicht werden um zusätzliche Mustererkennung, Varianz-Analyse, etc. zu ermöglichen. Die HES und PAAS Systemprogramme bestanden aus jeweils etwa 80 Routinen und Subroutinen, die in COBOL, FORTRAN 66 und Assembler Language geschrieben waren. Neben Update- und File-Maintenance Routinen hatten die Entwickler Programme zur Abfrage unter bestimmten Aspekten, zur Erstellung von Reports und geographischen Plots, sowie zur Analyse ausgewählter und frei kombinierbarer Attribute geschrieben. Die meisten dieser Programme sind in den Dokumenten umfassend beschrieben, durch Flussdiagramme abgebildet, und in Nutzerhandbüchern detailliert beschrieben.

Die Anwendung von experimentellem Reenactment zur Analyse von historischen Systemen der Informationsverarbeitung wirft einige praktische und ontologische Probleme auf, welche die Aussagekraft der Simulation für die Interpretation der historischen Rolle und Wirkung der Systeme beeinträchtigen. Augenscheinlich handelt es sich bei der kritischen Simulation um eine weitere Stufe der Modellierung und Abstraktion. Versuchten schon die Akteure, komplexe soziale Dynamiken in vereinfachte Modelle zu fassen, um sie im Medium der Datenbanken kalkulierbar und prozessierbar zu machen, so repliziert eine formale Modellierung des Systems diese Abstrahierung von den konkreten, historischen

Gegebenheiten ein weiteres mal. Es ist daher stets zu bedenken, dass durch die Simulation keine Aussagen über die Validität des Systems in der Repräsentation der Realwelt getroffen werden können, sondern nur über die Kohärenz der Informationen in ihrer Sammlung und Verarbeitung innerhalb des Systems. Zudem Abstrahiert eine Simulation und Replikation einzelner Prozessierungsschritte stark von den tatsächlichen materiellen und technischen Charakteristika des historischen Systems. So fand die Kommunikation mit dem System hauptsächlich mit Lochkarten statt, die sowohl für den Input von Daten als auch für das Laden von Programmen genutzt wurden. Wird diese materielle Dimension dieses historischen Mediums nur abstrakt modelliert oder gar ganz unterschlagen, so geraten maßgebliche Systemeigenschaften und Probleme, wie etwa der auf jeweils 80 Spalten beschränkte Speicherplatz oder die aufwändige und arbeitsintensive Normalisierung und Validierung der Daten, aus dem Blick. Gleiches gilt für Simulation der Verarbeitungsgeschwindigkeit und die Materialität der Maschinen und Prozessoren. Die IBM S/360 Systeme der späten 1960er Jahre waren raumfüllenden und extrem wartungsintensive Maschinen, die oftmals nicht physisch integriert waren, sondern aus Einzelkomponenten und Peripheriegeräten bestanden, zwischen denen Input und Output in Form von Lochkarten, die von Menschen hin- und her getragen wurden, ausgetauscht wurde. Die Berechnungszeiten selbst für einfache Operationen innerhalb solcher Systeme konnten Tage und Wochen beanspruchen, was ebenfalls bei einer Simulation in moderner Software nicht vergessen gehen darf. Schließlich ist die Dokumentation in zweifacher Hinsicht lückenhaft: Zum einen sind zwar die meisten der eingesetzten Programme und Subroutinen einigermaßen umfassend dokumentiert, aber das gilt eben nicht für alle und der Source-Code selbst ist nicht erhalten. Zum anderen handelt es sich bei Handbüchern um Darstellungen von einfachen und optimierten Nutzungsszenarien. Sie geben naturgemäß wenig Auskunft über Probleme und Fehler einzelner Anwendungen oder des Gesamtsystems. Zwar kann eine formale Modellierung Probleme mit der Datenkohärenz sichtbar machen, aber Probleme, die sich im Betrieb und in konkreten Situationen ergaben, bleiben bei dieser Analyse unentdeckt und lassen sich nur durch den Abgleich mit der historischen Dokumentation und eines möglichst breiten Quellenkorpus erfassen. Alle Simulationen können daher genau genommen nur als experimentelle Annäherungen gelten und müssen dringend umfassend dokumentiert, kritisch evaluiert und mit anderen historischen Zeugnissen abgeglichen werden.

4 Conclusion

Der Vietnamkrieg ist in die Geschichte eingegangen als krasser Fall der Mißinformation der Öffentlichkeit durch eine Regierung. Weder die politischen Motive, noch das militärische Vorgehen oder die tatsächliche strategisch-politische Situation wurden der amerikanischen und internationalen Öffentlichkeit durch die US-Regierung mitgeteilt, sondern kamen erst sukzessive, dank Whistleblowern, Leaks, mutige journalistische Arbeit und einer kritischen historischen Aufarbeitung ans Licht. Dabei ist das Fallbeispiel des Vietnamkriegs kein isolierter historischer Vorfall, denn sowohl die Konstellation - ein Bündnis aus westlichen Alliierten versucht in einer Gesellschaft der Dritten Welt in einen politischen Konflikt

einzugreifen - als auch die angewandten Mittel, nämlich eine Kombination aus militärischer Repression und politischer Manipulation auf Grundlage automatisiert erhobener und verarbeiteter Daten, sind auch in der Gegenwart und der Zukunft zu erwarten. Umso dringlicher ist die Frage, inwieweit die Fehlinformation der Öffentlichkeit eine bewusste Täuschung war und inwieweit sie auf einer fehlerhaften Interpretation der Fakten oder gar einer verzerrten oder gänzlich falschen Faktenlage beruhte. Um dies zu klären ist es unerlässlich, die Systeme der Informationsverarbeitung selbst einer historischen Analyse und Kritik zu unterziehen. Die Geschichtswissenschaft hat für den Umgang mit digitalen Objekten, seien es Datenbestände oder Systeme, bislang keine allgemein anerkannten analytischen Verfahren und Standards der Dokumentation entwickelt. Das in diesem Paper vorgestellte Experiment soll dazu beitragen, solche Standards und Verfahren zu entwickeln und zu prüfen. Es gründet auf der Annahme, dass eine Austausch und ein gegenseitiges Lernen zwischen den Geistes- und Sozialwissenschaften auf der einen Seite und den Informations- und Computerwissenschaften auf der anderen Seite für beide Seiten bereichernd sein kann und letztlich dazu beitragen wird, die Gesellschaft als Ganzes in die Lage zu versetzen, digitale Informationsverarbeitung und darauf beruhendes staatliches Handeln besser verstehen und kritisch bewerten zu können.

Literaturverzeichnis

- [B117] Blewer, Ashley: The PREFORMA handbook: validating formats, a prerequisite for preserving digital objects. PREFORMA, 2017. OCLC: 1263230909.
- [Bo21] Digital Preservation Coalition: Born digital archive cataloguing and description. Bericht, Digital Preservation Coalition, Mai 2021.
- [Ch21] Chakraborty, Indranil: Invisible labour: support service workers in India's information technology industry. Routledge, Abingdon, Oxon ; New York, NY, 2021.
- [Do17] Dourish, Paul: The stuff of bits: an essay on the materialities of information. The MIT Press, Cambridge, Massachusetts, 2017.
- [Du20] Dupré, Sven; Harris, Anna; Kursell, Julia; Lulof, Patricia; Stols-Witlox, Maartje, Hrsg. Reconstruction, Replication and Re-enactment in the Humanities and Social Sciences. Amsterdam University Press, November 2020.
- [Fe22] Feichtinger, Moritz J.: Computing Counterinsurgency: The Hamlet Evaluation System (HES) and Databasing During the Vietnam War. IEEE Annals of the History of Computing, 44(4): S. 28–43, Oktober 2022.
- [Fi21] Fickers, Andreas: Authenticity: Historical Data Integrity and the Layered Materiality of Digital Objects. In (Balbi, Gabriele; Ribeiro, Nelson; Schafer, Valérie; Schwarzenegger, Christian, Hrsg.): Digital Roots, S. 299–312. De Gruyter, August 2021.
- [FPS16] Fors, Hjalmar; Principe, Lawrence M.; Sibum, H. Otto: From the Library to the Laboratory and Back Again: Experiment as a Tool for Historians of Science. Ambix, 63(2): S. 85–97, April 2016.

- [FS76] Fry, James; Sibley, Edgar: Evolution of Data-Base Management Systems. *ACM Comput. Surv.*, 8: S. 7–42, März 1976.
- [Ha09] Haigh, Thomas: How Data Got its Base: Information Storage Software in the 1950s and 1960s. *IEEE Annals of the History of Computing*, 31(4): S. 6–25, Oktober 2009.
- [He20] Hendriksen, Marieke M. A.: Rethinking Performative Methods in the History of Science. *Berichte zur Wissenschaftsgeschichte*, 43(3): S. 313–322, September 2020.
- [Hi18] Hicks, Marie: Programmed inequality: how Britain discarded women technologists and lost its edge in computing. *History of computing*. MIT Press, Cambridge, MA London, UK, 2018. OCLC: on1004074349.
- [Ki14] Kitchin, Rob: The data revolution: big data, open data, data infrastructures & their consequences. SAGE Publications, Los Angeles, California, 2014. OCLC: ocn871211376.
- [Lo14] Loebel, Jens-Martin: Lost in translation: Leistungsfähigkeit, Einsatz und Grenzen von Emulatoren bei der Langzeitbewahrung digitaler multimedialer Objekte am Beispiel von Computerspielen. *Game studies*. Hülsbusch, Glückstadt, 2014. OCLC: 912044941.
- [Mc20] McIlwain, Charlton D.: Black software: the Internet and racial justice, from the AfroNet to Black Lives Matter. Oxford University Press, New York, NY, 2020.
- [Ri22] Ries, Thorsten: Digital history and born-digital archives: the importance of forensic methods. *Journal of the British Academy*, 10: S. 157–185, 2022.