Digital Humanities und Informatik: Lösungen für gemeinsame Probleme?

Vorname Name[1]

Institution[1]

Zusammenfassung

Bisher zeichnet sich der überwiegende Teil der Forschung in den historischen und Geisteswissenschaften durch mehr oder weniger qualitative Fragestellungen aus, die aus informatischer Sicht «unpräzise» und daher mit üblichen Lösungsansätzen kaum computerbasiert bewältigbar erscheinen mögen. Dies beginnt nicht erst mit unsicheren Datierungen und Zuschreibungen und endet noch längst nicht mit immer wieder neuen, bisherige Sichtweisen erweiternde Interpretationen von Texten und Artefakten. Zugleich haben diese Wissenschaften es mit Zeiträumen zu tun, die diejenigen in der Informatik um Größenordnungen überschreiten: Nicht nur für die (digital[isiert]en) Artefakte sondern auch und gerade für die Forschungsergebnisse in diesen Digital Humanities werden Aufbewahrungs-, Verfügbarkeits- und Bearbeitbarkeitszyklen in Jahrhunderten gemessen. Natürlich können die Digital Humanities für dieses informationstechnische Problem keine *technischen* Lösungen vorschlagen, aber ihre lange Tradition in der Aufbewahrung, Ordnung und Filterung von Informationen (im weitesten Sinne) könnte bei der Bewältigung dieses Problems hilfreich sein, indem sie strukturelle Ansätze für Grundmodelle und Verfahren bietet.

1 Die Problemlagen

Der *Call for Papers* für diesen Workshop hebt vor allem zwei Bereiche hervor, in denen sich Digital Humanities und Informatik möglicherweise ergänzen und gegenseitig bereichern können: Die «starke Heterogenität und Ambiguität» geisteswissenschaftlicher Daten und Forschungsgegenstände könnten einerseits als «spannende neue Herausforderungen und Betätigungsfelder» für die Informatik angesehen werden, während zugleich die «Geisteswissenschaften zumeist keine ausgeprägte Tradition hinsichtlich der quantitativ-empirischen Analyse großer Datenmengen» haben.

Ein drittes, im CfP — m.E. typischerweise — nicht angesprochenes, aber weit dringenderes Problem ist die langfristige Sicherung und Verfügbarkeit von Daten und den technischen Werkzeugen (Hard- wie Software) zu ihrer Nutzung und Weiterverarbeitung.

1.1 Heterogenität und Ambiguität von Forschungsdaten

Zu ersterem lässt sich vermuten, dass die Erfassung großer Datenmengen und ihre — ggf. KI-basierte — Indexierung in Zukunft hilfreich sein könnte, Beziehungen und Kontexte zu identifizieren, die einzelnen Geisteswissenschaftlern und selbst Forschungsgruppen mangels Übersicht über die riesigen Datenmengen bisher notwendig verborgen bleiben mussten und i.d.R. nur durch Zufallsfunde identifiziert werden konnten. Dies beginnt bei digitalisierten historischen Daten oder diesbezüglichen Forschungstexten bspw. schon mit der Ambiguität von Namensschreibweisen oder Begriffsbedeutungen in verschiedenen Sprachen und ihren historischen Stadien. Hier lässt sich vermuten, dass Ansätze wie «fuzzy logic» oder die maschinenbasierte Deutung von Kontexten und Bedeutungen durch Wortfeldanalysen der umgebenden Texte in Zukunft besonders hilfreich sein dürften: Bereits seit einigen Jahren schlagen Internetsuchmaschinen dem Nutzer auch Treffer für seine Suchanfragen vor, die naheliegende Themenbereiche oder ähnlichlautende Namen und Begriffe betreffen. Natürlich ist dies noch weit entfernt von der Präzision und Interpretationskomplexität (aber auch den Irrtümern), zu der menschliche Leser fähig sind, aber die Fortschritte bspw. bei sprachgesteuerten Assistenten lassen vermuten, dass hier in wenigen Jahren bereits deutlich bessere Ergebnisse erreichbar sind — selbst wenn sie «nur» zum Auffinden zusätzlicher Quellen und Texte oder Übersetzungen und Interpretationen führen, wofür Geisteswissenschaftler bisher bisweilen Monate oder gar Jahre in Archiven und Bibliotheken suchen mussten.

1.2 Große Datenmengen

Dies führt fast direkt zum zweiten Aufgabenkomplex, der sich zumindest in den meisten historischen Wissenschaften damit erklären lässt, dass bisher schlicht die technischen Möglichkeiten zur Erfassung der prinzipiell verfügbaren Datenmengen noch nicht gegeben waren, obwohl deren Bedeutung für die jeweilige Forschung i.d.R. längst erkannt wurde. Aber bisher und noch auf absehbare Zeit erscheint bspw. die vollständige digitale Erfassung historischer Archive und deren maschinenbasierte Aufbereitung für die wissenschaftliche Auswertung weitgehend illusorisch und selbst dort, wo theoretisch und technisch bereits möglich, schlichtweg nicht finanzierbar. Aber auch hier lässt sich erwarten, dass im Rahmen des rasanten technischen Fortschritts schon bald Lösungen verfügbar sein werden, die bspw. das schonende und schnelle Scannen von Archivbeständen sowie deren vorsortierende Interpretation erlauben.

1.3 Langfristige Datenverfügbarkeit

Im Unterschied zu den beiden vorgenannten Themen lässt sich für dieses Problem noch keine konkrete Lösung absehen, die über einen Zeitraum von mehr als 50 Jahren absehbar und sogar garantiert funktionieren würde. «If you have photographs you really care about, print them out», brachte Vinton Cerf dieses Problem seit 2015 pointiert zur Sprache: Als TCP/IP-Entwickler einer der «Väter des Internet» und in seinen Funktionen als Vizepräsident bei Google sowie als Präsident der ACM dürfte Cerf sicherlich mit den technischen Entwicklungen und ihren Fehlstellen gut genug vertraut sein, um einschätzen zu können, ob solche Lösungen

bisher existieren oder wie sie aussehen könnten bzw. müssten. Ähnliches gilt für Alan Kay, den Mitentwickler objektorientierter Programmierung und graphischer Benutzeroberflächen.

2 Lösungsansätze

Im Folgenden sollen zu allen drei oben genannten Probleme absehbare oder vorhandene Lösungsansätze skizziert und diskutiert werden. Während für die beiden ersten allein die technische Entwicklung und eine engere Kooperation zwischen Digital Humanities und Informatik sowie deren (bisher nicht) ausreichende Finanzierung als gangbarer Lösungsweg erscheint, stellt das dritte insbesondere die Informatik vor Herausforderungen, für welche die Geisteswissenschaften zwar keine technischen, wohl aber möglicherweise strukturelle Lösungsansätze bieten können.

2.1 Heterogenität und Ambiguität – ein Beispiel

Im Jahr 2001 begann der Verfasser, gemeinsam mit dem Kunsthistoriker Dr. Martin Raspe und unterstützt von einem Projektteam an der Bibliotheca Hertziana, dem Max-Planck-Institut für Kunstgeschichte in Rom, mit der Entwicklung einer Datenbank für die Architekturzeichnungen des römischen Barock unter dem vom Renaissance-Theoretiker und -Architekten Leon Battista Alberti entliehenen Titel *Lineamenta*.[1] Basierend auf dem damals neuen, freien Web Application Server ZOPE[2] und der damit verbundenen, objektorientierten Datenbank ZODB[3] wurde für eine erste Testversion eine hierarchische Struktur geschaffen, die bereits vorhandene Module des überwiegend in Python[4] geschriebenen Anwendung nutzte, um die noch nicht implementierten Funktionen eines Content Management Systems für die interaktive, webbasierte Nutzung der Datenbank bereit zu stellen. Kurze Zeit später wurden diese allerdings in einer für ZOPE standardisierten Form mittels des ZOPE-eigenen «Content Management Frameworks» (CMF)[5] verfügbar, auf dem dann neben anderen, ähnlichen CMS-Lösungen das wiederum kurz darauf verfügbare, heute noch intensiv weiter entwickelte und im Vergleich zu anderen, bekannteren Lösungen nicht nur sicherere sondern auch weiterhin Objektorientierung bevorzugt unterstützende CMS «Plone»[6] beruhte. Zugleich entstand in der weltweiten Entwickler-Community um ZOPE ein Framework, dass die graphische Modellierung der (objektorientierten) Klassen, ihrer Beziehungen untereinander sowie ihrer Eigenschaften und

[1] Vgl. http://lineamenta.biblhertz.it und die dort verlinkten Dokumente zum Projekt.

[2] Vgl. http://www.zope.org

[3] Vgl. http://www.zodb.org

[4] Vgl. https://www.python.org

[5] Vgl. http://old.zope.org/Products/CMF/index.html/

[6] Vgl. https://plone.org

Methoden mittels UML[7] erlaubte. Daraufhin entstand in der römischen Arbeitsgruppe die Idee, das angestrebte, weit komplexere Datenmodell für eine Forschungsdatenbank, die über die Testversion von *Lineamenta* weit hinaus gehen sollte, den Entwicklern bei einem *Sprint* (2003 im Vorarlberg) zu präsentieren, um so ggf. professionelle Unterstützung bei der Umsetzung zu erhalten. Tatsächlich stieß das äußerst komplexe Datenmodell mit seinen vielfach (noch) nicht eindeutig zu definierenden Klassen und Methoden auf großes Interesse seitens der Entwickler, denn es war abzusehen, dass die Lösung *dieses* komplexen Problems einen «Werkzeugkasten» bereithalten würde, der für viele kommerzielle, jedoch i.d.R. weit weniger anspruchsvolle Anwendungen in Wirtschaft und Wissenschaft leicht anwendbar und leicht anpassbar sein würde. Die Problem- und Fragestellungen einer historischen Wissenschaft schienen also geeignet, eine Entwicklung zu initiieren, die unmittelbar praktische Anwendbarkeit und Nutzen in gänzlich anders gearteten Aufgabenfeldern versprach. Leider erwies sich das in Rom entwickelte Datenmodell als so komplex, dass es in einem gemeinsamen *Sprint* mit führenden Entwicklern der ZOPE-Community nicht «zum Laufen» gebracht werden konnte. Jedoch lässt sich konstatieren, dass die (auf den ersten Blick) negativen Ergebnisse dieser Anstrengungen dazu geführt haben, dass einige der Entwickler sich das Schweizer Datenmodell GEVER[8] für die elektronische Datenverarbeitung in der Administration auf allen hierarchischen Ebenen der Verwaltung — vom Gemeindearchiv bis zur Bundesverwaltung der Schweiz — zum Vorbild nahmen und diese auch erfolgreich in der ZOPE-basierten Lösung «OneGov GEVER»[9] umsetzten (was namhaften CMS-Herstellern zuvor *nicht* gelungen war).

Es erscheint durchaus im Bereich des Möglichen, dass heute ein erneuter Versuch der Umsetzung des *Lineamenta*-Projekt- und Datenmodells eher gelingen könnte, zumal vor allem die Rechenkapazitäten seit 2003 enorm gewachsen sind. Allerdings erheben sich hier anwendungsbezogene Bedenken: Während der Arbeiten am und Diskussionen zum Projekt *Lineamenta* und dem «erweiterten Datenbankbaukasten für geisteswissenschaftliche Projekt» ZUCCARO[10] zeigte sich zunehmend, dass eine fertige Lösung aufgrund der hohen Komplexität (es wurden allein über 120 Objektklassen im Modell definiert!) kaum noch von «normalen» Anwendern benutzbar sein würden: Die Einarbeitung in die aktive Bentutzung des Systems (Eingabe neuer und Ergänzung oder Verknüpfung vorhandener Daten) wäre so langwierig und die Arbeit mit dem System von so viel Spezialwissen abhängig gewesen, dass eine Einbeziehung möglichst vieler Wissenschaftler nicht realistisch erschien — schon gar nicht in der von mir befürworteten Form eines offenen Wikis für namentlich registrierte Benutzer.

---

[7] https://de.wikipedia.org/wiki/Unified_Modeling_Language

[8] Vgl. https://de.wikipedia.org/wiki/GEVER und https://www.isb.admin.ch/isb/de/home/themen/bundesarchitektur/schwerpunkte/geschaeftsverwaltung_gever.html

[9] Vgl. https://onegovgever.ch

[10] Akronym für «Zope-based Universal Configurable Classes for Architectural/Arthistorical Research Online»; vgl. http://zuccaro.biblhertz.it

Im späteren Verlauf wurde nach dem Ende meines befristeten Vertrags 2004 zuerst eine Umstellung auf die XML-Datenbank eXist-db[11] vorgenommen und später das vollständig anders strukturierte CIDOC-CRM[12] zugrunde gelegt: Während das ursprüngliche Datenmodell davon ausging, den nicht interpretationsabhängigen *Ort* eines Artefakts zum Ausgangspunkt der (hierarchischen) Objektdatenbankstruktur zu nehmen, weil so den wissenschaftlichen Nutzern ein «natürlicher» Zugang wie bspw. zu einem Archiv möglich wäre: ausgehend vom Ort (ggf. noch gruppiert in einer Oberklasse «Land») über Institution, Sammlung, Teilsammlung, Konvolut, Blatt, Blattseite, Einzelzeichnung sollte die Hierarchie flexibel den Zugang zum (Daten-) Objekt ermöglichen, benutzt CIDOC-CRM bekanntlich das *Ereignis* als die grundlegende Ordnungskategorie, wobei mir nie einleuchtete, wie man eine schwer datierbare Zeichnung eines ggf. unbekannten Zeichners eines über lange Zeiträume entstandenen Bauwerks wie St. Peter in Rom *einem* Ereignis zuordnen soll…

Damals gelangte ich zu der bisher immer wieder bestätigten (jedoch von der Leitung der Projekte in Rom abgelehnten) Auffassung, dass man derartige Projekte nach dem «KISS»-Prinzip[13] gestalten sollte, um möglichst viele der im Grund ja nur vergleichsweise wenigen Wissenschaftlerkollegen in die Sammlung und Auswertung der Daten einbeziehen zu können und so die (immer viel zu geringen) finanziellen Mittel möglichst effektiv einzusetzen. Gleichzeitig sollte eine zu enge Fokussierung der zu erfassenden Daten auf eine bestimmte Forschungsfrage (oder eine sehr spezifische Gruppe solcher Fragen) durch «Überbestimmtheit» der zu erfassenden Daten vermieden werden, weil dies deren Nachnutzung für andere Projekte erschwert.

Insgesamt bin ich aber aufgrund dieser und ähnlicher Erfahrungen überzeugt, dass es von großem Nutzen für beide Seiten sein kann, wenn komplexe und «unscharfe» Fragestellungen aus den historischen und Geisteswissenschaften als «spannende» und «herausfordernde» Aufgaben an die Informatik herangetragen werden. Ich sähe hier z.B. eine breite Möglichkeit, Studierende beider Fächer nicht nur in der Kommunikation miteinander, der sie im Berufsleben sicherlich noch häufiger begegnen werden, zu «trainieren», sondern auch, interessante Aufgaben bspw. für Studienprojekte, Abschlussarbeiten u.ä. zu finden bzw. zu definieren.

2.2 Große Datenmengen

Wie erwähnt, haben auch die historischen und Geisteswissenschaften es oftmals mit sehr großen Datenmengen zu tun, selbst wenn die riesigen Archive bspw. Venedigs, der Medici oder des Vatikans im Vergleich zu den Datenmengen, die bei Industrie und Banken anfallen, «klein» erscheinen mögen. Aber selbst deren Sichtung und Auswertung war bisher kaum umfassend möglich, sondern konzentrierte sich auf ganz bestimmte Archivalien oder — wo deren Quellen auffind- und überschaubar waren — konkrete Fragestellungen. Das Hauptproblem scheint dabei bisher vor allem die fehlende «manpower» zu sein, die allein bspw. für das

[11] Vgl. http://exist-db.org/exist/apps/homepage/index.html

[12] Vgl. http://www.cidoc-crm.org

[13] Vgl. https://de.wikipedia.org/wiki/KISS-Prinzip

Einscannen oder Transkribieren von Quellen nötig ist: Die «Fingerschatten» oder fehlenden ausklappbaren Tafeln in den Scans historischer Bücher, die bspw. vom Google-Projekt gescannt wurden, zeigen, dass eine befriedigende technische Lösung allein für das korrekte Umblättern und ggf. Ausklappen von Seiten in Büchern oder von losen Blättern in Archiven noch nicht in Sichtweite zu sein scheint. Setzen wir aber einmal voraus, dass Roboter dazu schon bald in der Lage sein werden, so wäre im zweiten Schritt die Erkennung und die darauf basierende Zuordnung der erfassten Daten mittels OCR und struktureller Ordnungsschemata ein weiteres, bisher kaum gelöstes Problem, das sich aber im Übrigen mit dem der «Heterogenität und Ambiguität» der Daten überschneidet.

Die historisch entwickelten Ordnungsschemata wie alphabetische, systematische und Standortkalatoge sowie bspw. die Dewey-Dezimalklassifikation in Bibliotheken oder die Findbücher in Archiven mögen aus informationstechnischer Sicht als rudimentäre Vorläufer heutiger Datenbankkataloge erscheinen, man sollte dabei aber berücksichtigen, dass sie immer nur «Gedankenstützen» für den Bibliothekar, Archivar und auch den Benutzer bildeten, derren komplexes und weiter reichendes Wissen unverzichtbar für die (auch «kreative») Deutung der in handgeschriebenen Büchern oder auf Karteikarten festgehaltenen Daten war und ist. *Dieses* Wissen in Computersystemen abzubilden scheint mir noch die weit komplexere Herausforderung für die Bewältigung großer Datenmengen in der Zukunft zu sein, wobei insbesondere das Ausschließen von *false positives* eine interessante Herausforderung darstellen dürfte: Ein «Niederschlag» dieses Problems ist in der Filterung der Ergebnisse selbst der präzisesten Anfrage an Internet-Suchmaschinen oder Bibliotheksdatenbanken zu erkennen, die letztlich immer noch den menschlichen Nutzer verlangt, ohne dessen zusätzliche Kenntnisse die Interpretation und Auswahl der Ergebnisse (noch) unzureichend bleiben muss. Es ist aber zu erwarten, dass insbesondere maschinelles (oft noch von Menschen trainiertes) Lernen hier jedoch schnell Fortschritte erzielen wird.

2.3 Langfristige Datenverfügbarkeit

Das Problem der langfristigen Datenverfügbarkeit lässt sich in zwei Teilprobleme gliedern: Zum einen und zuallererst besteht es in der — im Vergleich zu historischen Objekten und ihren wissenschaftlichen Interpretationen — geradezu grotesk kurzen «Lebensspanne» informationstechnischer Lösungen sowohl in der Hard- wie in der Software, von denen wohl keine bisher länger als 30, maximal 50 Jahre existiert hat, ohne dass die Daten — wenn überhaupt — unter erheblichem Aufwand transferiert werden mussten… und von denen ebenso sicher wohl auch keine länger als 50 Jahre Bestand haben wird. Zwar werden z.Z. für Sicherung wissenschaftlicher Datenmengen Containerformate vorgeschlagen, das «Forschungsdatenmanagement» erscheint bereits 28 Jahre nach der Vorstellung des explizit für den wissenschaftlichen Austausch entwickelten WWW — und damit also mindestens 25 Jahre zu spät — auf der Agenda von Wissenschaftsfördereinrichtungen und Ministerien. Aber die Entwicklung der letzten Jahrzehnte lässt befürchten, dass die Wiederherstellung dieser Daten und der zu ihrer vollständigen, erhebenskonformen Nutzung notwendigen Soft- und ggf. Hardware kostspieliger und langwieriger werden wird als ihre Neuerhebung mit dann jeweils aktueller Technik, der auf lange Sicht natürlich das gleiche Schicksal droht.

Vinton Cerf schlägt hierfür als Lösung ein *Digital Vellum* genanntes System vor[14], das nicht nur die Datenformate und die zu ihrer Bearbeitung notwendige Software, sondern auch die für deren Betrieb notwendigen Betriebssysteme und notfalls (also sicher!) sogar die Hardware emulieren können soll. M.W. ist die Entwicklung eines solchen Systems nach mehreren Jahren noch nicht abgeschlossen; seine freie Verfügbarkeit steht wohl in den Sternen und die für die Speicherung der bereits heute anfallenden Daten notwendigen Speichermedien wächst währenddessen exponentiell. (Von lizenz- und datenschutzrechtlichen Problemen bei der Wiederherstellung und Nutzung «historischer» Daten ganz abgesehen…)

Alan Kays Ansatz, gemeinsam entwickelt mit Long Tien Nguyen, lässt schon im Namen erkennen, dass er einen buchstäblich «konservativeren» Weg verfolgt: «Cuneiform Tablets of 2015»,[15] also «Keilschrifttafeln von 2015» stellen selbstbeschreibende Speichermedien dar, deren Finder selbst in tausenden von Jahren bei hinreichender Intelligenz und technischer Kapazität in der Lage sein soll, die Maschine zu bauen, die zum Auslesen der Daten notwendig ist. Diese sind dann zwar — eben ähnlich wie Keilschrifttafeln für den geschulten Wissenschaftler — (halbwegs) lesbar, hoffentlich auch verständlich, jedoch als *read only memory* nicht direkt nutz- und z.B. beschreibbar. D.h., die mit ihnen gespeicherten Daten müssen erst wieder in dann aktuelle Systeme übertragen werden, um sie nutzen zu können. Auch hier dürfte in vielen Fällen die Frage nach Aufwand und Nutzen von unseren Nachfahren leider negativ beantwortet werden…

Beide Lösungsvorschläge stehen zudem vor dem zweiten Teilproblem: Wie sollen die zu erhaltenden Daten ausgewählt werden? Denn die anfallenden Daten übersteigen inzwischen bereits jetzt bei weitem die Speichermöglichkeiten — wenn man einmal von der angeblichen Kapazität des neuen Speicherzentrums der NSA in Utah absieht, das in der Lage sein soll, den *gesamten* Internet-Verkehr von drei oder sogar mehr Jahre zu speichern … zumal eine Anfrage an die NSA nach den Daten eines vielleicht vor mehreren Jahren eingestellten Forschungsprojekts eher negativ beschieden werden dürfte.

Inwieweit könnten die historischen und Geisteswissenschaften hier hilfreich sein? Man ist versucht zu antworten, dass sie bspw. seit Jahrhunderten über Kriterien für die Auswahl zu speichernder Daten verfügen. Allerdings wäre dagegen kritisch einzuwenden, dass es gerade die ehemals (und oft bis vor kurzem) für überflüssig und nicht aufbewahrenswert gehaltenen Daten sind, die interessante Forschungsfragen aufwerfen oder bei der Beantwortung alter wie neuer Fragen von zuvor unschätzbarem Nutzen sind. Daraus wäre also abzuleiten, dass man lieber mehr als zuwenig Daten aufheben sollte … nur ist das eben angesichts des rasanten Wachstums der Datenmengen wohl auf lange Sicht nicht praktikabel, weil sowohl die Speichermedien als auch die Indexierungs- und Katalogisierungstechniken nicht ausreichen. Beispielsweise ist es gängige Praxis in der Wirtschaft wie in der Verwaltung, Akten über finanzielle Vorgänge und ihr digitales Pendant nur wenige Jahre aufzuheben, in der Regel so lange, wie die Steuergesetzgebung es verlangt. Aber gerade Abrechnungen bspw. für Kunstwerke oder auch nur Löhne helfen uns oft nach Jahrhunderten als einzige, Werke zu datieren.

[14] Vgl. z.B. https://www.nitrd.gov/news/digital-vellum-and-archives.aspx

[15] Vgl. http://www.vpri.org/pdf/tr2015004_cuneiform.pdf

Direkte, also sozusagen «materiale» Kriterien für die Aufbewahrung von Daten lassen sich so m.E. nicht gewinnen. Natürlich kann man einwenden, dass zumindest die Daten aus Forschungsprojekten (im Unterschied zu privaten oder privatwirtschaftlichen Daten) aufzuheben seien: Dies mag für historische und geisteswissenschaftliche Projekte (noch) möglich sein, das CERN oder astrophysikalische Forschungsprojekte stoßen jedoch schon längst an Speichergrenzen und verarbeiten daher gesammelte Daten nur noch im *streaming*, d.h. sie werden möglichst direkt nach dem Erfassen unter bestimmten Kriterien ausgewertet und dann gelöscht. D.h. sie stehen zukünftig nicht mehr zur Verfügung und müssten ggf. neu erhoben werden.

Allerdings bin ich der Auffassung, dass sich aus den historischen *Erfahrungen* im Umgang mit (ehemals) großen Datenmengen wie den o.g. Archiven, wie sie eben in den historischen Wissenschaften vorliegen, nicht materiale, aber *strukturelle* Kriterien für die langfristige Datenspeicherung und *darauf basierend* vielleicht sogar für die geeignetsten technischen Lösungen ableiten lassen, die von Seiten der eher kurzfristig denkenden und handelnden privat(wirtschaftlich)en Datenerzeugung und -nutzung nicht zu erwarten sind. Die historische Entwicklung der schriftlichen Aufbewahrung von Daten scheint dafür zu sprechen, dass admininstrative *und künstlerische* (!) Anforderungen die technische Entwicklung der Speichermedien voran getrieben hat: Sie reicht von den Inschriften in Stein und Bronze (z.B. wurden wichtige Gesetze so im römischen Tabularium am Forum Romanum aufbewahrt) über Keilschrifttafeln aus Ton, weiter über Papyrusrollen und Pergamentcodices aus einzelne Seiten bis zum papiernen Buch, das zuerst von Hand beschrieben und später gedruckt wurde.

Es könnte also bspw. eine lohnenswerte Aufgabe wissenschaftlicher Kooperation zwischen Digital Humanities und Informatik sein, historische Prozesse, Speichermedien und -kriterien zu analysieren und als heuristisches Modell für neue technische Entwicklungen zu nehmen, die das bisher ungelöste Problem der langfristigen Datensicherung lösen helfen könnten.

Nur als abschließendes Beispiel: Es wäre m.E. denkbar, das softwaretechnische Modell von Alan Kays *Cuneiform tablets* mit Vinton Cerfs *Digital Vellum* zu verschmelzen: Ersteres beruht (nicht sehr verwunderlich) auf dem Ansatz, der bereits *Smalltalk* zugrunde lag, an dessen Entwicklung Kay führend beteiligt war: Eine komplexe Arbeitsumgebung inklusive aller Funktionen wie Programmierung, Speicherung, sogar Datenbanken, sollte in einer virtuelle Maschine laufen, deren Verbindung zur und damit Anforderung an die Hardware so minimal wie möglich definiert und realisiert werden konnte und wurde. Im Ergebnis waren die Dialekte von *Smalltalk* m.W. wohl auf bis 120 verschiedenen Computerarchitekturen nutzbar und benötigten in vielen Fällen dank hardware-naher Programmierung der *virtuelle Maschine* nicht einmal ein Betriebssystem. Auch wenn es «hart» klingen mag, so halte ich es für überlegenswert, die Entwicklung eines solchen Systems neu, *from scratch* und unter quasi «zentraler» Leitung in Angriff zu nehemn (vergleichbar den Nationalbibliotheken des 19. Jahrhunderts, die als Reaktion auf die Bücherflug entstanden), in das dann die jeweilige Anwendungssoftware portiert oder vergleichbares neu geschrieben werden könnte. Die zentrale Institution, die ähnlich einer Zentralbibliothek oder einem Nationalmuseum mit «Ewigkeitsgarantie» ausgestattet sein müsste (und nicht mit maximal dreijährige Projektstellen aus einem Sondertopf für «Forschungsdatenmanagement…»), könnte die Entwicklung koordinieren, ihre langfristige Stabilität und die Übernahme aller Daten und Software garantieren — etwas, was Vinton Cerfs *Digital Vellum* meiner Einschätzung nach nicht leisten können wird.