

# WIE MAN DAS VERSCHWINDEN UNSERER DATEN IM "DIGITALEN SCHWARZEN LOCH" UND EIN "DUNKLES INFORMATIONENZEITALTER" VERHINDERN KÖNNTE

Bernd Kulawik

be\_kul@posteo.de

**KURZDARSTELLUNG:** Der Beitrag versteht sich als Fortsetzung und Ergänzung desjenigen zur Vorjahrestagung, der unter dem Titel "Digitales Kuratieren – und dann?" ein düsteres Bild von der Zukunft aller digitalen (nicht nur: Forschungs-) Daten zeichnete, welches noch über die pessimistischen Prognosen Vinton Cerfs hinaus ging, da weder das von Cerf vorgeschlagene "digital vellum" noch das – wohl eher nur sarkastisch von ihm vorgeschlagene – simple Ausdrucken aller Dokumente als Lösungsansätze überzeugen können. Im Folgenden wird die Auffassung vertreten, dass (nicht nur) die Wissenschaft durch die zunehmende Digitalisierung aller Lebensbereiche und also auch Forschungstätigkeiten vor einer strukturell ähnlichen, jedoch noch weit größeren Herausforderung steht, als die Bibliotheken des 19. Jahrhunderts angesichts der explodierenden Produktion an Druckerzeugnissen: Der damalige Lösungsansatz, diesem Problem durch die Gründung von Nationalbibliotheken und die Schaffung bibliothekarischer Standards zu begegnen, welche *alle* Druckerzeugnisse sammeln und ihre Inhalte such- und findbar machen sollten, wird als Vorbild für den Umgang mit digitalen Daten angesehen. Im Unterschied zu erst seit kurzem – also ca. 25 Jahre zu spät – laufenden Bestrebungen, das 'Forschungsdatenmanagement' zu vereinheitlichen und dezentral mit Blick auf langfristige Speicherlösungen zu organisieren, wird hier ein Ansatz vorgeschlagen, der strukturell demjenigen des 19. Jahrhunderts ähnelt: Die Schaffung einer mit 'Ewigkeitsgarantie' ausgestatteten Institution, die nicht nur eine freie, auf die größtmögliche Zahl von Hardwaresystemen übertragbare Softwareumgebung für die Forschung bereit stellt, sondern auch deren standardisierte (Weiter-) Entwicklung betreibt bzw. kontrolliert sowie alle damit erzeugten Daten z.B. nach Abschluss von Forschungsprojekten in ein Repositorium übernimmt und so deren nicht nur passive, sondern auch aktive Verfügbarkeit für die Dauer der eigenen Existenz garantiert.

## 1. EINFÜHRUNG

Im vergangenen Jahr habe ich an dieser Stelle in den schwärzesten Farben die Zukunft aller Daten ausgemalt, die wir in den sog. *Digital Humanities* tagtäglich produzieren und womit nicht nur Texte, sondern Datenbankeninhalte, angepasste Software, dieser zugrundeliegende Betriebssysteme und eigentlich auch noch die Hardware gemeint sind. Nach Vinton Cerf werden sie in einem "information black hole" verschwinden – und unsere Zeit wird zukünftig das "digital dark age" genannt werden. [1] Dabei ging ich sogar noch über die finsternen Prognosen Cerfs hinaus, die sich buchstäblich bereits tagtäglich erfüllen, die aber auf einem Paradigma für den Umgang mit Computern stehen, welches durch die Verwendung von Metaphern wie Dokument, Datei, Ordner,

Desktop, Dateisystem, (Einzelplatz-) Rechner usw. charakterisiert werden kann. Denn genau dieses Paradigma löst sich gerade vor unseren Augen bspw. in der 'Cloud' und der 'Timeline' auf, welche den Umgang der jüngeren Generation mit Computern prägen.

Natürlich ist das zugrundeliegende Problem schon länger bekannt, mindestens seitdem günstigere Computer und Speichermedien sowie das Internet in Gestalt des WWW eine größere Verbreitung erfahren – also seit mindestens ca. 25 Jahren. Und seit 2–3 Jahren wird bereits sogar über Lösungen nachgedacht, die über den Ratschlag hinaus gehen, möglichst alle Dateien regelmäßig auf je aktuelle Speichermedien zu kopieren und mittels der aktuellsten Version der Erstellungssoftware in die aktuellste Version des Dateiformats zu

überführen. (Sie tun dies doch hoffentlich regelmäßig z.B. mit Ihren alten WordPerfect-Dateien...?)

## 2. LÖSUNGSVORSCHLÄGE

Im Folgenden möchte ich kurz, aber hoffentlich nicht verfälschend, einige Lösungsansätze für das oben (und ausführlicher im vorigen Jahr) skizzierte Problem vorstellen und einige m.E. entscheidende Kritikpunkte nennen.

### 2.1 AUSDRUCKEN

Wie weit die bisherigen Lösungsansätze allgemein sind, zeigt indirekt Cerfs Rat, alle digitalen Dokumente sicherheitshalber auszudrucken... [1] Selbst wenn *alle* Daten, die im Rahmen wissenschaftlicher Forschung gesammelt, bearbeitet oder als Ergebnisse publiziert werden, in Form druckbarer Dokumente vorliegen würden: Dabei gingen gerade die *Verknüpfungen* der Informationen, aus denen m.E. nicht nur wissenschaftliches Wissen wesentlich besteht, verloren bzw. sie wären wieder bestenfalls als Fussnoten darstellbar. D.h., wir würden alle Vorteile, welche die Arbeit mit Computern gegenüber derjenigen mit Federkiel, Tinte und Bleisatz hat, über Bord werfen müssen. Dies kann wohl kaum ein befriedigender Lösungsansatz sein. Dass Cerf ihn trotzdem mehrfach und sicherlich nicht nur sarkastisch präsentiert hat, kann wohl nur so gedeutet werden, dass die Lage aus seiner Sicht sehr ernst ist: Und einer der 'Väter des Internet', der zugleich als Vizepräsident bei Google und Präsident der ACM, der ältesten und größten im IT-Bereich tätigen Berufsorganisation sicherlich über den Stand der aktuellen technischen Entwicklungen sowie ihre Möglichkeiten und Beschränkungen im Bilde sein sollte, dürfte wissen, wovon er hier redet bzw. wovon er damit warnt.

### 2.2 DIGITAL VELLUM

Vint Cerf hat seit 2014 als Lösung für das Problem des drohenden Datenverlustes ein sog. *Digital Vellum* ("digitales Pergament") vorgeschlagen [2], das nach meinem Verständnis im Wesentlichen in einer IT-Umgebung besteht, die das ständige Übertragen der Daten übernehmen oder eher sogar ersetzen soll, indem sie 'einfach' alle zur weiteren Nutzung der Daten notwendigen Elemente in sich aufnimmt, nachbildet, emuliert bzw. virtualisiert: Dateiformate, Anwendersoftware, Betriebssystem und sogar Hardware. Allerdings scheint es, dass

ausgerechnet der Erfinder des TCP/IP die zunehmende Vernetzung solcher Daten nicht nur in Intranets – z.B. zum Bilddatenbankserver eines Instituts –, sondern auch über verschiedenste Server im Internet, die nicht unter Kontrolle des Urhebers eines spezifischen Datensatzes/Dokuments stehen und auch nicht ewig verfügbar oder erreichbar sein werden, dabei noch gar nicht berücksichtigt hat. Anderenfalls müsste das System wohl eine 'Kopie des Internets' und aller angeschlossenen Rechner umfassen und diese fortlaufend aktualisieren. Damit dürfte selbst die NSA überfordert sein...

### 2.3 CUNEIFORM TABLETS OF 2015

Long Tien Nguyen und Alan Kay haben 2015 eine Lösung vorgestellt [3], die weniger die fortlaufende Nutzung digitaler Daten als ihre mögliche Wiederherstellung selbst noch nach Jahrtausenden anstrebt. Man könnte sie überspitzt als ein *digital vellum* beschreiben, das seine eigene Beschreibung und eine Anleitung zu seiner Wiederherstellung ausgehend von einfachsten mathematischen Konzepten selbst enthält und somit sogar einem außerirdischen Archäologen der fernsten Zukunft erlauben soll, unsere Datenträger – soweit sie erhalten wären – (a) überhaupt als solche zu identifizieren und (b) die Maschine zu ihrer Benutzung zu konstruieren. In einem wesentlichen Punkt unterscheidet sich dieses System jedoch m.E. vom *digital vellum*: Die Autoren schlagen eine Einbettung der Daten samt der gesamten notwendigen Software (also inklusive der ggf. zu emulierenden Hardware) in eine Virtuelle Maschine vor, die sich auf minimale Anforderungen und einen entsprechend reduzierten Satz an Befehlen beschränken soll und so jederzeit auf nahezu jeder im Moment vorstell- oder absehbaren Hardware schnell rekonstruierbar wäre. Dieses Konzept erinnert natürlich nicht zufällig an die Sprache *Smalltalk*, an deren Entwicklung Kay um 1970 beteiligt war (s.u.). Wie einleitend erwähnt, scheint diese Lösung aber für den absehbaren bzw. wünschbaren täglichen Gebrauch und die entsprechend zu fordernde interaktive Nutzung der Daten, also ihre fortgesetzte Änderung, nicht wirklich geeignet.

### 2.4 FORSCHUNGSDATENMANAGEMENT (DEUTSCHLAND, 2016)

In Deutschland gibt es unter dem Schlagwort Forschungsdatenmanagement [4] z.Z. eine Flut von Ausschreibungen, die anscheinend alle auf

eine Förderinitiative des BMBF zurück gehen und mit denen – natürlich nur für befristete Anstellungen! – Personen gefunden werden sollen, die nicht nur in dem jeweiligen Fachgebiet der ausschreibenden Projekte und Institutionen einigermaßen bewandert, sondern auch in der Lage sind, eine Lösung für das Problem der langfristigen Verfügbarkeit von Forschungsdaten zu entwickeln. Ob dieser Ansatz tauglich sein mag, wage ich nicht zu beurteilen, allerdings kann man sich schon fragen, wie sinnvoll es ist, für ein auf 'Ewigkeit' angelegtes Vorhaben, viele kleine, voneinander unabhängige und kurzzeitig befristete Projektstellen zu schaffen. Ich fürchte, dass sich damit die durchschnittliche (aufgrund eigener Erfahrungen geschätzte) Verfügbarkeit der in einem üblicherweise befristeten Projekt erzeugten Daten von vielleicht zehn auf maximal zwanzig Jahre erhöht. In der Regel ist diese Verfügbarkeit anscheinend an die Anstellungsdauer der involvierten Mitarbeiter oder ihrer Kollegen, die zukünftig ein Interesse an diesen Daten haben, gekoppelt: Die nächste Generation wird den Aufwand zur Rettung der Daten vermutlich bereits für größer halten als den daraus zu ziehenden Nutzen: Und so sind denn die Projekte in den *Digital Humanities*, die bereits wieder 'gestorben' sind und deren Daten sich gar nicht oder nur mit riesigem Aufwand auf museumsreifer Hardware wiederherstellen ließen, bereits Legion.

### **2.5 DATA MANAGEMENT PLAN (SCHWEIZ, 2017)**

In der Schweiz geht man einen etwas anderen Weg: Man stellt nicht eigens Mittel für die befristete Anstellung von Fachwissenschaftlern als 'Hobby-Programmierern' bereit, sondern der Schweizerische Nationalfonds (Äquivalent zur DFG) verlangt seit Kurzem zu jedem Projektgesuch einen *Data Management Plan* [5]: D.h., antragstellende Wissenschaftler sollen selbst einen überzeugenden Plan vorlegen – sich also das notwendige IT-Wissen aneignen oder entsprechend Hilfe suchen –, mit welcher Software welche Daten erhoben und wie diese in Backups gesichert sowie langfristig verfügbar gehalten werden sollen. Der SNF möchte also anscheinend, dass sich Wissenschaftler selbst zu Spezialisten heranbilden, die ein Problem lösen, für welches die *Computer Sciences* selbst noch *keine* Lösung haben...

### **2.6 RFII 2016**

Inzwischen hat sich in Deutschland der *Rat für Informationsinfrastrukturen* gebildet, dessen Namen vielleicht die Älteren an eine um 2000 aufgelegte Förderlinie der DFG erinnert, die den schönen Namen *Informationsinfrastrukturen für netzbasierte Forschungskooperation und digitale Publikation* trug. M.W. wurden die damals bereitgestellten Mittel nicht oder nur in so geringem Umfang abgerufen, dass man diese Initiative wieder abbrach. Der RfII schlägt nun, gut 15 Jahre später – und damit wiederum ca. 25 Jahre zu spät – vor, die Bemühungen zum Forschungsdatenmanagement im weitesten Sinne zu bündeln und auf wissenschaftspolitischer sowie wissenschaftlicher Ebene eine breite Diskussion und Kooperation in Gang zu setzen, welche die Ziele und die Rahmenbedingungen zu ihrer Erreichung definieren und *in the long run* in einer nationalen Forschungsdateninfrastruktur (NFDI) realisieren soll: "Dieses soll als bundesweites, verteiltes und wachsendes Netzwerk arbeitsteilig angelegt sein." [6]

Jemand, der in verschiedenen europäischen Ländern, sogar 'außerhalb' der EU, gearbeitet und dabei mit den bürokratischen 'Monstern' föderaler 'Forschungsförderung' (die man korrekterweise 'Forschungs- und Lebenszeitvernichtung' nennen sollte) gerungen hat, kann diesen Vorschlag nur mit Skepsis lesen. Aber vorausgesetzt, dieses Vorhaben ließe sich realisieren (woran angesichts des Wildwuchses an 'Standards' und jeweils projektbezogener Eigenentwicklungen gezweifelt werden darf, was aber trotzdem irgendwie angesichts der skizzierten Lage gelingen *muss*): Die Betonung des 'Nationalen' bereits im Titel der 'NFDI' macht schaudern. Sie ist als vermutliche Reaktion auf die erwähnten 'Monster' der EU-Bürokratie sicherlich verständlich, angesichts der notwendig internationalen Verschränkung heutiger Forschung aber nicht akzeptabel, sondern ein Schritt zurück.

### **3. ALTERNATIVVORSCHLAG**

Es dürfte deutlich geworden sein, dass ich keine der skizzierten Lösungen für das Problem der tatsächlich *langfristigen* Datensicherung *und* -verfügbarkeit als ausreichend erachte.

#### **3.1 'DIGITALE ZENTRALBIBLIOTHEK'**

M.E. befinden wir uns in einer ähnlichen Situation wie die Bibliotheken im 19. Jahr-

hundert, als die Zahl der Publikationen buchstäblich explodierte: Es war schnell klar, dass keine der damals z.T. seit Jahrhunderten existierenden Bibliotheken diese Menge vollständig aufnehmen und ihre Erschließung würde leisten können. Die Schlussfolgerung war die Schaffung der Nationalbibliotheken, die mit entsprechenden Budgets und vor allem einer 'Ewigkeitsgarantie' ausgestattet, dies leisten sollten und bisher auch sehr gut geleistet haben. Nimmt man die – nicht mehr häufig zu hörende – metaphorische Rede vom Internet bzw. der Digitalisierung als einer Erweiterung der zweidimensionalen *Gutenberg-Galaxis* um eine dritte oder gar vierte Dimension jedoch ernst – und dafür spricht m.E. Vieles –, so kann die Schlussfolgerung nur lauten, dass es heute einer ähnlichen Institution bedarf, um die digitalen Daten nicht nur zu retten sondern eben auch zu erschließen und so (nicht nur wissenschaftlichen) Interessenten überhaupt dauerhaft zur Verfügung stellen zu können. Es böte sich zumindest an, eine solche Institution als eine europaweite zu starten, sie jedoch für alle anderen Interessenten offen zu halten, wobei *Europa* hier explizit nicht – wie leider in jüngster meist üblich – nur als die EU verstanden werden darf, denn gerade die (Kultur-) Geschichte Europas ist ein fast unentwirrbares Knäuel an Relationen, und die Artefakte, die es dokumentieren, und deren Erforschung und Bewahrung *dürfen natürlich nicht* kurz- und durchsichtigen tagespolitischen Interessen untergeordnet, also geopfert werden. Die Bindung einer solchen Institution an die EU erschiene sogar vor dem Hintergrund gerade der Politik der letzten Jahre als *design to fail*.

### 3.2 STRUKTUR

Diese vorgeschlagene Institution sollte natürlich – und kann m.E. auch nur – dezentral in dem Sinne aufgebaut sein, dass sie über Niederlassungen an verschiedenen Orten verfügt, die das Gesamtsystem zugleich als verteilte, parallele Datenspeicherorte gegen den Verlust einer einzelnen 'Kopie' immunisieren: Ein Gedanke, der bekanntlich bereits dem Arpanet, dem Vorläufer des Internet, zugrunde lag. An diesen Standorten könnten zugleich Wissenschaftszentren angesiedelt sein, die für Tagungen oder Kooperations-treffen inklusive Unterbringung der Teilnehmer genutzt werden könnten, damit der unsägliche und absurd teure, die immer zu knappen Forschungsmittel verschwendende Tagungs-

tourismus endlich reduziert oder sogar überwunden werden kann. Eine optimale Verbindung dieser Zentren durch nichtkommerzielle Hochleistungsverbindungen könnte weiter zu einer erheblichen Reduktion der ja auch nicht umweltfreundlichen Reisetätigkeit durch Video-Konferenzen beitragen. Ein modulares, also jederzeit veränderbares Gebäudesystem für solche wissenschaftlichen Arbeitszentren wurde bereits 1997/98 für die Schweizer Landesausstellung *Expo.01* vom Architekten und Designer Fritz Haller – von dem das *USM Haller* Stahlbaumöbelsystem stammt – unter dem Titel *Haus der Zukunft* vorgeschlagen. [7]

### 3.3 SOFTWARE

In Analogie zu den Standards und Werkzeugen, welche die Nationalbibliotheken des 19. Jahrhunderts entwickelten – i.d.R. Katalogsysteme und Bibliographien –, sollte die hier skizzierte 'Digitale Zentralbibliothek' eine eigene, freie und quelloffene Software professionell entwickeln und bereitstellen, welche *alle* Bedürfnisse der Forschung abzudecken in der Lage sein muss. Das wäre also ungefähr das, was DFG und SNF von den Wissenschaftlern verlangen... Das mag auf den ersten Blick ein grotesk umfangreiches und nie zu erreichendes Projekt sein. Aber zum einen gibt es das, verteilt auf einige große, meist US-amerikanische Firmen sowie eine Vielzahl kleiner, freier Softwareprojekte schon – d.h., man müsste solche Firmen und Initiativen 'einfach' nur einbinden bzw. auf wenige, klare, einfache aber hart durchzusetzende, offene Standards verpflichten. Diese Software sollte für alle, die öffentliche Fördergelder in Anspruch nehmen oder einfach nur ihre Daten dieser Einrichtung zur 'ewigen' Aufbewahrung anvertrauen möchten, verpflichtend sein. Natürlich ist der Aufschrei aus den *scientific communities* schon jetzt zu erahnen, dieser Zentralismus würde die Freiheit der Forschung zerstören. Aber es gibt ihn längst: Seine Zentralen heißen Redmond, Cupertino, Redwood City oder Walldorf – nur sind sie am Gewinn interessiert, nicht an der möglichst langfristigen Nutzbarkeit ihrer Produkte und der damit erzeugten (Forschungs-) Daten! Und zum anderen lautet die Alternative zur Entwicklung einer solchen Software: Verlust *aller* nicht ausgedruckten Daten in spätestens 50 und der gedruckten in vielleicht 200 Jahren. Angesichts der öffentlichen Gelder, die bisher über Forschungsprojekte zumeist in private Taschen geflossen sind, sowie der Arbeits- und

also Lebenszeit von Millionen Forschenden darf das keine akzeptable Perspektive im Sinne eines *Ruhig weiter so!* sein.

### 3.4 STRUKTUR DER SOFWTARE

Es klang schon an, dass die zentral bereitstellende Softwarelösung möglicherweise vollständig neu zu schreiben wäre: Der Wildwuchs der letzten Jahrzehnte hat zu so vielen, scheinbar parallelen aber i.d.R. zueinander (nicht zuletzt aus Gründen der Gewinnmaximierung) inkompatiblen Lösungen geführt, dass der Versuch, sie zusammen zu führen (wie man ihm beim *RfII* herauslesen kann), eigentlich zum Scheitern verurteilt sein muss. Ausserdem böte ein *reboot* oder ein *start from scratch* die Chance, nicht nur viele bekannte Fehler und Mängel in älterer, zum 'Industrie'-Standard gewordener Software sowie dazugehörigen Protokolle und Formaten zu vermeiden, sondern diese auch so zu konzipieren, dass ihre zukünftige Veränderung und Anpassung an gänzlich andere, neue, heute noch notwendig unbekannte Aufgaben und Systeme jederzeit möglich wäre. Natürlich kann man so etwas auch heute nur nach 'bestem Wissen und Gewissen' auf dem aktuellen und absehbaren Stand der Technik tun – und wir dürfen zwar annehmen, dass die Entwickler der Software, die wir heute nutzen, sich zumindest bemüht haben, ähnlich vorzugehen. Aber wir *wissen*, dass dies nicht immer der Fall war, dass z.B. bestimmte Protokolle oder Formate 'mit der heißen Nadel gestrickt' wurden oder grundsätzliche Designentscheidungen – man denke nur an die absichtliche Inkompatibilität der Windows-Pfadbeschreibung mittels \ im Unterschied zu dem in den Unices verbreiteten / – dem Gewinnstreben untergeordnet wurden.

Für die Struktur der Software und ihrer Verteilung stelle ich mir ein 'geschlossenes' System vor, in das nicht autorisierte Software bspw. gar nicht erst eindringen kann (was nebenbei die Sicherheit erhöhte): So könnte man anstelle der etablierten, aber wohl nur aus Gründen der Hardwarekosten und für der knappen Ressourcen einmal präferierte von-Neumann-Architektur, die Daten und Operationen im selben Speicherbereich ausführt, die Harvard-Architektur o.ä. bevorzugen. Dieses System sollte mit der notwendigen Hardware nur über eine möglichst einfache und kleine *Virtuelle Maschine* (VM) kommunizieren, so dass ihre Portabilität auf alle bekannte Hardware jederzeit leicht sichergestellt werden könnte. Alle Daten, Programme, Dateisystem

sowie Kommunikationskanäle und -protokolle wären in diese VM eingebettet. Sie müsste so entworfen sein, dass sie problemlos vom Handheld Computer über den – zunehmend bereits unbeliebten – Arbeitsplatz-PC bis zum Server und der ggf. sogar weltweit verteilten Serverfarm (Cloud) problemlos skaliert. Linux tut das im Prinzip ja bereits. Mit der rasant fortschreitenden Verfügbarkeit von Rechenkapazität selbst in kleinsten Geräten und eigentlich unbegrenzt verfügbarer (erneuerbarer) Energie sollte das längst möglich sein.

Zur Aufbewahrung und weiteren Nutzung bereits mit älterer Software erzeugter Daten, deren Transfer in die Office-Programme, Datenbanken etc. des neuen Systems nicht oder nur mit unvertretbarem Aufwand möglich wäre, könnten in diesem wiederum Virtualisierungen bereitgestellt werden.

Dieses System müsste nicht nur für alle steuerfinanzierten Forschungsprojekte verpflichtend, sondern auch kostenfrei sein: Jeder, der damit privat oder außerhalb der Forschung arbeiten möchte, sollte es frei nutzen können. Trotz der notwendig zu verlangenden Quell-offenheit und Freiheit (bezogen auf Lizenzen, Patente u.ä.) dürften aber Änderungen an der Software bspw. zur Datenerfassung für spezifische Forschungsaufgaben selbst nur in Rücksprache mit der 'Zentralbibliothek' oder sogar nur von dieser vorgenommen werden, um 'Wildwuchs' zu vermeiden: Wie in der freien Software üblich, könnte natürlich jeder nach Belieben seinen *fork* des Gesamtsystems erstellen –, wäre für diesen dann aber auch selbst verantwortlich, eine Verantwortung für die damit erzeugten Daten, die zumindest über Jahrzehnte niemand ernsthaft übernehmen wollen dürfte bzw. könnte. — Die Verteilung und der Einsatz der Software könnte so organisiert sein, dass *Images* oder '*virtual boxes*', also vollständig funktionsfähige Instanzen des Systems sowie jene Anwendungssoftware, die für ein Projekt benötigt wird, vom Nutzer heruntergeladen werden. Nach Abschluss des Projekts oder auch für das tägliche Backup könnten die erzeugten Daten dieses Subsystems jederzeit wieder in das der 'Zentralbibliothek' hochgeladen bzw. auf deren Server 'gespiegelt' werden. Damit könnten zugleich, bei Vorliegen einer entsprechenden, mit 'sanftem' Druck eingeforderte Freigabe der Daten durch die Urheber, *Open Access* und *Open Data* umfassend gewährleistet werden.

Die Einbindung in das übergreifende System der 'Zentralbibliothek' könnte zugleich zumindest für alle Daten, die ebenfalls mit diesem erzeugt wurden und darin früher oder später enthalten sind, die Stabilität von Verlinkungen garantieren – ein heute nicht oder kaum lösbares Problem, obwohl es unstrittig sein dürfte, dass *Wissen* im Verknüpfen von *Informationen* (Daten) besteht.

### 3.5 HARDWARE

"Leute, die ernsthaft an Software interessiert sind, sollten ihre eigene Hardware bauen", hat der für treffsichere Formulierungen bekannte Alan Kay einmal gesagt. [8] Aufgrund des bisher Ausgeführten liegt es nahe, dass eigentlich auch die Hardware, auf der die zu entwickelnde Software (optimal) läuft, möglicherweise neu entwickelt und den Anforderungen der Software angepasst werden sollte. Hierfür wären wiederum ebenfalls ausschließlich freie, offene Designs und Bauelemente zu nutzen, so dass jeder Hersteller diese in *freier* Lizenz bauen, aber kein Monopol damit erzielen könnte. Auch bei der Konzeption dieser Hardware ließen sich sicherlich eine Menge Fehlentscheidungen, Irrtümer und vor allem überholte, aber aus Kompatibilitätsgründen immer noch 'mitgeschleppte' Mängel vermeiden (man denke an das A20 Gate). Ihre Entwicklung und Herstellung dürfte zudem nicht nur ein Konjunkturmotor werden, sondern könnte zugleich jene Nachhaltigkeit durchsetzen, die von kommerziellen Herstellern bisher natürlich aus Gründen der Gewinnmaximierung vermieden werden muss.

## 4. WAS KÖNNTE MAN ÄNDERN?

Was könnte man im Zuge der Entwicklung eines solchen neuen Systems *from scratch* anders machen? Nur einige Gedanken dazu:

### 4.1 SICHERHEIT

Grundsätzlich ist es sicherlich unumgänglich, dass die Sicherheit der Daten, der Zugriffsberechtigungen und des ganzen Systems nach den besten heute verfügbaren Konzepten von Grund auf in das System integriert, aber nicht nur deren Implementierungen, sondern sogar die Konzepte selbst zugleich jederzeit austauschbar sein müssten – Stichwort 'Modularität'. Es ist ja bereits jetzt absehbar, dass alle bisher verfügbaren oder denkbaren Sicherheitssysteme in naher Zukunft durch Quantencomputer zumindest leicht umgangen werden können. Blockchain-artige Konzepte scheinen

hier aktuell am geeignetsten, wenn auch ihre Berechnung noch riesige Ressourcen benötigt.

### 4.2 SMALLTALK 1972 / SQUEAK 1996

Von der 'Programmiersprache' *Smalltalk* haben Sie sicher schon einmal gehört, aber vermutlich nicht, dass sie eigentlich eine graphische 'Entwicklungsumgebung' – und zwar die früheste dieser Art – mit einer eigenen 'virtuellen Maschine' darstellt. Tatsächlich läuft diese geniale Software auf so ziemlich allem, was mit Nullen und Einsen rechnen kann – oder sie lässt sich dank der sehr kleinen VM relativ leicht auf neue Hardware übertragen. Von *Smalltalk* gibt es eine Implementierung "written in itself" namens *Squeak*, in der also eine Version dieser *Smalltalk*-Umgebung selbst wiederum in *Smalltalk* geschrieben wurde. Dies hat z.B. den Vorteil, dass man im laufenden Betrieb alle Teile des Systems verändern kann – inklusive der VM selbst! *Squeak* wurde von Alan Kay, einem der an *Smalltalk* Beteiligten, entwickelt, um Kinder 'jeden Alters' spielerisch an das Programmieren heran zu führen. Es sollte also den schon sehr einfachen Zugang zu *Smalltalk* – der Name ist Programm! – weiter vereinfachen und tat dies auch: Über Jahrzehnte schrieben tausende Kinder hunderttausende kleine und größere Anwendungen in *Squeak*, die frei verfügbar waren. Kennen Sie eine Programmiersprache und -umgebung, von der sich Ähnliches sagen ließe? – Die Konzepte von *Squeak* leben heute in *E-Toys* und *Scratch* weiter und sind für die Heranführung von Kindern ans Programmieren und an das Verstehen von Computern immer noch hervorragend geeignet. Wenn Sie sich darüber informieren wollen, was in der IT in den letzten 50 Jahren falsch gelaufen ist, schauen Sie sich eine der vielen Präsentationen Alan Kays im Internet an oder lesen Sie seine Texte bzw. Interviews.

Falls Ihnen die oben kurz beschriebenen Konzepte z.B. von Java her bekannt vorkommt: Ja, sie sind ähnlich, nur wurde *Smalltalk* gut 20 Jahre früher entwickelt – in der IT-Branche eigentlich mehr als ein Zeitalter –, ist portabler und kann mehr. Aber man kann das Rad ja mehrmals 'neu' erfinden, und sogar schlechter als das Original: Eine 'Legende' aus der 'Steinzeit' des Computerzeitalters berichten, dass einst ein junger, computerbegeisterter Unternehmer im Xerox PARC (Palo Alto Research Center) *Smalltalk* sah, wo es für eigene Entwicklungen entstan-

den war. Er erhielt aber von Xerox keine Lizenz, es zu vermarkten, weil man es dort gar nicht als ein vermarktbare Produkt ansah. Der junge Mann ließ daraufhin von Freunden und den Angestellten seines Startups etwas Ähnliches nachprogrammieren, das aber entscheidende Eigenschaften nicht hatte. Sein Name war Steve Jobs; sein Startup hieß Apple. – Ein anderer junger Mann wurde später beschuldigt, die Ideen für die graphische Benutzeroberfläche seines Betriebssystems von Apple gestohlen zu haben, konnte aber das Gericht überzeugen, dass er sie ebenfalls bei einem (etwas späteren) Besuch im Xerox PARC kennen gelernt und direkt übernommen (also an der Quelle und nicht beim ersten Kopisten 'gestohlen') und in seiner Firma diese Ideen und Konzepte, so gut es eben ging und soweit man sie verstanden hatte, ebenfalls hatte nachprogrammieren lassen. Dieser junge Mann hieß Bill Gates... Ich bin überzeugt, dass *Squeak/Smalltalk* zumindest strukturell hervorragend als Modell, wenn nicht sogar als *starting point* für die Entwicklung der oben skizzierten Software dienen könnte. Zumindest ist mir keine bessere Alternative bekannt...

### 4.3 ENGELBARTS PRÄSENTATION 1968

Sie kennen vielleicht Douglas Engelbarts berühmte Präsentation von 1968, oft "the mother of all presentations" genannt. [9] Wenn Sie sie nicht kennen, lassen Sie alles stehen und liegen, schauen Sie sie sich an, am besten mehrmals hintereinander, und achten Sie genau auf *alle* Elemente – die visuellen wie den Text: Und dann fragen Sie sich, welche der demonstrierten Eigenschaften Ihr heutiger, viele Millionen mal schnellerer Computer *nicht* beherrscht – Achtung Spoiler: Es sind fast alle! Wenn Sie das Gesehene und Gehörte verstanden haben, wird es Ihr Leben vermutlich verändern und Ihre Achtung vor den 'Errungenschaften' der Computerindustrie in den letzten 50 Jahren ... modifizieren. Vielleicht überlegen Sie sich auch noch, wie Ihre tägliche Arbeit aussähe, wenn Ihr Computer diese Möglichkeiten – und sei es nur als Alternative zum Gewohnten – anböte. Zum Beispiel startet Engelbart zur 'Textverarbeitung' kein eigenes Programm, sondern schreibt direkt auf dem 'Desktop'. Er springt ständig zwischen dem, was wir als Textprogramm und 'Dateimanager' getrennt kennen und nutzen würden, hin und her, ohne dabei entsprechende 'Programme' und deren 'Fenster' zu öffnen, zu schließen oder zwischen ihnen zu wechseln.

Wenn er zwischen verschiedenen Ansichten desselben 'Dokuments' oder einer Liste von 'Dokumenten' wechseln will, tut er dies ebenfalls ohne Änderungen über ein Menü u.ä. — usw. usf. [Falls Sie sich über Engelbart und diese Präsentation genauer informieren wollen sollten – und das sollten Sie –, lesen Sie *bitte nicht* Henning Lubins Buch *Engelbarts Traum* von 2015: Es steckt so voller Irrtümer, Fehler und Behauptungen, dass man es eigentlich Satz für Satz widerlegen müsste...]

Engelbarts Präsentation hält jedenfalls so viele Konzepte, Ideen und 1968 (!) bereits vorliegende Lösungen bereit, dass einige davon für eine – für jede! – neu zu entwickelnde Software in Betracht gezogen werden sollten.

### 4.4 FTW.BOOK

Die Berner Firma *4teamwork* [10] hat 2015 unter der Bezeichnung *ftw.book* [11] für das freie Content Management System *Plone* ein Modul entwickelt, welches unter Betreuung des Verfassers im *Google Summer of Code 2016* [12] etwas weiter entwickelt wurde. *Plone* basiert auf dem ebenfalls freien, in der Skriptsprache *Python* geschriebenen Application Server *Zope*, der über eine eigene, objektorientierte Datenbank verfügt. Das Modul, *ftw.book*, erlaubt es, in *Plones* Webinterface neue Templates für Buch, Kapitel und Textseite(n) zu nutzen, die für den Nutzer als normale, hierarchisch gestaffelte Webseiten erscheinen und – ebenso wie andere Inhalte dieses CMS – durch Suchmaschinen indizierbar sind sowie im WYSIWYG-Modus bearbeitet werden können. *ftw.book* bietet aber zusätzlich über eine Anbindung an LaTeX die Möglichkeit, jederzeit aus den aktuellen Webseiten TeX- oder PDF-Dateien zu erzeugen und bspw. an ein *Print-on-demand*-System zu übergeben, welches aus diesen dann instantan ein gedrucktes Buch erzeugen kann. Alle an der Entstehung eines Textes Beteiligten können so nicht nur gemeinsam an *demselben* Textdokument samt Versionierungssystem arbeiten – jeder hat sicherlich schon die leidvolle Erfahrung gemacht, dass bei 'netzbasierter Forschungskooperation' unterschiedliche Versionen einer Datei zirkulierten und heillose Verwirrung stifteten –, und jeder kann jederzeit durch Ausgabe des PDFs das in dem Moment aktuelle, prinzipiell aber endgültige Layout sehen und kontrollieren. Mir ist keine Editions-Software bei großen Wissenschaftsverlagen bekannt, die Ähnliches bietet. Da *ftw.book* alle LaTeX-Klassen für wissen-

schaftliche Publikationen nutzen kann, ist die Verfügbarkeit professioneller Layouts ebenso gewährleistet, wie das (unabsichtliche) Zerstören derselben durch die Bearbeiter ausgeschlossen. (Versuchen Sie mal, dies mit dem auch für diesen Text genutzten Word-Template sicher zu stellen!) Für die Empfehlung Cerfs, vorläufig alle Dokumente auszudrucken, wäre dies also vielleicht eine gangbare Lösung.

## 5. SCHLUSS

Sollte der Eindruck entstanden sein, dass ich hier eine neue *Diktatur* zentralistischer Bevormundung für die Forschung vorschlagen wollte, so möchte ich dem schärfstens widersprechen: als 'gelernter DDR-Bürger' bin ich mir der Gefahren und Einschränkungen durch solche Systeme durchaus bewusst. Ich bin aber auch überzeugt, dass wir für die Lösungsfindung im 'ergebnisoffenen' Diskurs und unter Beteiligung aller *stake holder* (insbesondere der kommerziell ausgerichteten Industrie) nicht nur keine Zeit mehr haben, sondern es auch *stake holder* geben dürfte, die genau solch eine Lösung mit allen Mitteln zu verhindern suchen würden. (LiMux, *anyone?*)

Zugleich bin ich überzeugt, dass es nicht nur möglich sein sollte und wäre, in einem 'gesiteten' Diskurs jeglichen 'Machtmissbrauch' seitens einer solchen Institution zu vermeiden, sondern dass die zu fordernde vollständige Freiheit einer solchen Lösung unter Umständen auch dazu führen könnte, dass sich das Bessere in Form eines *forks* durchsetzt, falls die 'Zentrale' selbst diesen nicht zu unterstützen bereit ist (vgl. Open/Libre Office).

Ich bin überzeugt, dass alle, die aktuell wissenschaftliche Daten erzeugen oder nutzen ein massives Interesse daran haben und sich also dafür einsetzen *sollten*, dass eine dauerhafte und nachhaltige Lösung des eingangs beschriebenen Problems schnellstmöglich und ggf. unter Einsatz aller nötigen Mittel erreicht werden sollte, denn bis zu diesem Zeitpunkt werden nicht nur alle bisher eingesetzten Mittel verloren gehen, im "information black hole" Vinton Cerfs verschwinden, sondern vor allem auch Ihre Arbeit, Ihre Mühen, Ihre Daten und Erkenntnisse. Die vorgeschlagene Lösung muss ja nicht die richtige sein – aber vielleicht und hoffentlich provoziert sie produktiven Widerspruch, damit es endlich voran geht.

## 6. DANKSAGUNG

Bedanken möchte ich bei allen Genannten, aber insbesondere bei den Mitgliedern des Squeak e.V., die mir auf ihrer Jahrestagung 2017 am Hasso-Plattner-Institut Potsdam die Möglichkeit gaben, zentrale Fragen des hier vorgestellten Konzepts zu diskutieren.

## 7. LITERATURHINWEISE

Alle Links wurden letztmalig am 19. November 2017 überprüft.

- [1] <https://www.theguardian.com/technology/2015/feb/13/google-boss-warns-forgotten-century-email-photos-vint-cerf>
- [2] Vint Cerf auf der 25. Jahrestagung des W3C über das *digital vellum*: <https://vimeo.com/110794988> und ausführlicher: [www.youtube.com/watch?v=STeLOogWqWk](http://www.youtube.com/watch?v=STeLOogWqWk)
- [3] Nguyen, Long Tien; Kay, Alan: The Cuneiform Tablets of 2015. Viewpoints Research Institute, VPRI Technical Report TR-2015-004, Los Angeles: 2015, online [www.vpri.org/pdf/tr2015004\\_cuneiform.pdf](http://www.vpri.org/pdf/tr2015004_cuneiform.pdf)
- [4] <https://www.bmbf.de/foerderungen/bekanntmachung-1233.html>
- [5] [http://www.snf.ch/de/derSnf/forschungspolitische\\_positionen/open\\_research\\_data/Seiten/data-management-plan-dmp-leitlinien-fuer-forschende.aspx](http://www.snf.ch/de/derSnf/forschungspolitische_positionen/open_research_data/Seiten/data-management-plan-dmp-leitlinien-fuer-forschende.aspx)
- [6] RfII – Rat für Informationsinfrastrukturen: Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. Göttingen: 2016, S. 2.
- [7] "Haus der Zukunft", in: Stalder, Laurent; Vrachliotis, Georg (Hrsg.): Fritz Haller – Architekt und Forscher. Zürich: gta Verlag, 2015, S. 326-327.
- [8] <https://www.quora.com/People-who-are-really-serious-about-software-should-make-their-own-hardware-Why>
- [9] <http://web.stanford.edu/dept/SUL/library/extra4/sloan/mousesite/1968Demo.html>
- [10] <https://www.4teamwork.ch>
- [11] <https://onegov.ch/verein/community/zertifizierte-module/ftw.book>
- [12] <https://summerofcode.withgoogle.com/archive/2016/projects/6028139876581376>