

1. Guten Morgen und vielen Dank an die Organisatoren, dass ich hier heute sprechen darf. Ich bin kein Informatiker, jedenfalls nicht gemäss Studium und Hauptberuf, aber ich beschäftige mich seit 1982 mit Computern und ihrer Programmierung sowie ihrem Einsatz vor allem in den historischen Geisteswissenschaften. Denn ich habe — nach etwas Physik — Musikwissenschaft, Philosophie und Kunstgeschichte studiert und arbeite zur Zeit zu einem riesigen Bestand an Architekturzeichnungen aus der Renaissance, die vor allem antike Bauten dokumentieren.
2. Es geht dabei also um Daten auf 450 Jahre alten Objekten, die 2000 Jahre alte Objekte betreffen und zu denen seit über 150 Jahren geforscht und publiziert wird. Dies nur als Gedankenstütze für die Zeitrahmen, in denen ich arbeite und in denen wir bezüglich unserer eigenen Daten und Objekte und deren Publikation denken oder denken können sollten. . .

Das heisst, ich bin auch nicht im digitalen Publikationswesen tätig, sondern nutze es nur, wenn möglich, um selbst etwas zu publizieren – z.B. die Forschungsdaten aus meinen Projekten.

Heute morgen möchte ich Ihnen, wenn Sie gestatten, etwas Angst machen, indem ich den gegenwärtigen Zustand nicht nur des digitalen Publikationswesens und der langfristigen Forschungsdatensicherheit in den schwärzesten Farben male, sondern die Gegenwart und Zukunft unseres Umgangs mit Computern und ähnlichen Geräten sowie den damit verarbeiteten Daten. Dabei werde ich zwei Lösungsansätze von Leuten präsentieren, die nicht nur quasi seit der Steinzeit der heutigen Computer »dabei« sind und Entscheidendes zu ihrer Nutzung beigetragen haben, sondern immer noch sehr aktiv und daher wohl auch in der Lage sind, den Zustand so einschätzen zu können, dass man ihne zuhören sollte, wenn man mich mit gutem Recht nicht für kompetent genug hält. Allerdings möchte ich gleich anmerken, dass ich beide Lösungsansätze, deren Realisierung noch nicht abgeschlossen ist, nicht für ausreichend halte, da sie m.E. wesentliche Probleme nicht berücksichtigen. Damit wird also hoffentlich die Hoffnung, die ich wecken möchte, sofort wieder zerstört. Verstehen Sie mich bitte nicht falsch: Ich habe überhaupt keine Freude daran, den »advocatus diaboli« zu spielen. Aber ich denke, die Probleme sind so ernst, dass wir mit dem »worst case scenario« rechnen und für *dieses* eine radikale Lösung finden müssen. Es wird also eher die meiste Zeit nur am Rande um *Digitales Publizieren* im strengen Sinne gehen, sondern — wenn man so will — um die digitale Umwelt, in der das digitale Publizieren heute und auf absehbare Sicht ebenso stattfinden wird wie der Umgang mit Forschungsdaten.

3. Vor fast genau drei Jahren wirbelte Vinton Cerf, Vice President bei Google mit dem schönen Titel »Internet Evangelist« und Präsident der ACM, der Association for Computing Machinery, der 1947 gegründeten und also wohl ältesten und zugleich grössten wissenschaftlichen Gesellschaft für Informatik im weitesten Sinne einigen Staub auf, als er vor dem Verlust all unserer Daten in nur wenigen Jahrzehnten warnte:
4. Nahezu alle unsere Daten würden früher oder später, aber eher früher, in einem »information black hole« verschwinden. Unsere Zeit werde daher für die Zukunft als »digital dark age« erscheinen. Und sein Rat an alle Betroffenen lautete: »If there are photos you really care about, print them out.« Es leuchtet ein, dass diese Aufforderung auf *alle* digitalen Formate, nach der üblichen Metaphorik also: Dokumente im weitesten Sinne anzuwenden ist. (Verlage, die Informationen auf »totem Holz«, vulgo »Papier« verewigen und verbreiten, werden das gern hören. . .) — Falls Sie Cerf nicht glauben, so vergegenwärtigen Sie sich einfach, welche Speichermedien, Datenformate und Computer Sie als erstes benutzt haben, wie sich diese zu den heutigen verhalten und wie diese wohl in einer Zukunft aussehen, die genauso weit von heute entfernt ist wie Ihre ersten Erfahrungen. In meinem Fall sind es 35 Jahre: 1982 wurden meine ersten Programme noch über eine Tastatur mit 8 Tasten (und einer »Enter«-Taste) in eine Art selbst gebauten Computer im Maschinencode eingegeben und auf Lochstreifen aus Papier gespeichert. Langfristige Datensicherheit hätte damals bedeutet: Lagerung der Papierrollen in einem feuerfesten, wasserdichten Schrank. Ich bin überzeugt, dass wir *keinen Grund* zu der Annahme haben, die technische Entwicklung und damit die Veränderung der Medien, Datenformate, Hardware etc. werde in den nächsten 35 Jahren weniger schnell vonstatten gehen. Tatsächlich beschleunigt sie sich aber.
5. Cerf schlägt für die Lösung dieser Probleme ein System vor, dass er *Digital vellum*, also digitales Pergament, nennt. Grob gesagt soll es ein »Röntgenbild« eines älteren oder aktuellen kompletten IT-Systems in einer Art »Schnappschuss« konservieren und diese »Image« in einer virtualisierten Umgebung auf einer Hardware jederzeit bereit halten können, die selbst langfristig stabiler ist als unsere heutige.
6. Die Eigenschaften des *Digital Vellum* umreisst Cerf in dieser Vortragsseite von 2014 wie folgt:

Während sich die technischen Details eigentlich von selbst verstehen, finde ich besonders den vierten Punkt interessant: »special copyright provisions for

preservation practices«. Hat jemand einen Überblick über alle aktuell bzw. bis heute gültigen Copyright- bzw. Urheberrechtssysteme, deren Implementierung oder mögliche Formalisierung zwecks Berücksichtigung in einem zu programmierenden IT-System? Auch die Entwicklung formaler, einheitlicher Beschreibungen für Datenstrukturen ist ja schon seit längerem in Arbeit, »harzt« aber etwas, wie der Schweizer sagen würde: Während die Vorschläge Tim Berners-Lees für das WWW von 1991 binnen drei Jahren soweit umgesetzt waren, dass man es wirklich bereits weltweit benutzen konnte, bleibt die Entwicklung des ebenfalls von Berners-Lee vorgeschlagenen *Semantic Web* doch »etwas« hinter den Erwartungen: In wissenschaftlichen Bereichen wie der Wissenschaftsgeschichte oder dem wissenschaftlichen Bibliothekswesen wurde sogar schon explizit deren Scheitern behauptet, weil sich deren »Objektbereich« nicht in eine wie komplex auch immer geartete »Ontologie« pressen lassen würde. Wem dieser Begriff seit der Beschäftigung mit Sprachphilosophie ohnehin schon immer ein Graus war, der dürfte sich in der Annahme bestätigt sehen, dass es eigentlich keine »Ontologien« geben kann. . .

7. Natürlich sieht auch Cerf aufgrund seiner Erfahrung einige Herausforderungen für das *Digital Vellum*, die nicht nur oder weniger technischer Natur sind: Denn selbst wenn es gelingt, die Bits eines Word-Dokument, evtl. noch im alten, aber weiterhin gern genutzten binären .doc-Format, in 50 Jahren noch als ein digitales Dokument zu identifizieren und ggf. wiederherzustellen sowie die Hardware (oder deren Virtualisierung) samt Betriebssystem zum Laufen zu bringen: Spätestens wenn die im Snapshot enthaltene Word-Kopie beim Start die Gültigkeit des Lizenzschlüssels durch eine Nachfrage beim Microsoft-Server in Redmond überprüfen will, dürfte klar werden, dass uns bzw. zukünftige Nutzer unserer Daten hier einige Probleme erwarten.

Und wer meint, dass sich gerade an die vielfältigen Lizenzen, Patente, Urheber- und Publikationsrechte sowie die allgemeineren rechtlichen Rahmenbedingungen in 50 Jahren ohnehin niemand mehr erinnern werde, sei mit einem Blick auf die Debatten in Literatur und Musik daran erinnert, dass zumindest *Anwälte* diese Fragen – und die damit verbundenen Verdienstmöglichkeiten – ganz bestimmt *nicht* vergessen werden . . .

8. Ich hatte erwähnt und Sie wissen vielleicht ohnehin, dass Vint Cerf im IT-Bereich nicht irgendwer ist: Seine erwähnten Positionen verdankt er nicht zuletzt der Entwicklung des TCP/IP um 1973/74, welches wir heute noch munter benutzen, obwohl es vermutlich nie für all das gedacht war, was es heute ab-

decken muss. Aber immerhin hat es sich als erstaunlich robust erwiesen.
Cerfs Kollege bei der Entwicklung des TCP/IP war Robert Kahn,

9. der 2016 auf einer Konferenz in Bern auf weitere Probleme der langfristigen Datensicherung und -nutzung hinwies: Neben dieser Aufgabe, der digitalen Aufbewahrung des digitalen Erbes der Welt, und mit ihr engstens verbunden
10. sind nämlich nicht nur technische (und einige nicht vernachlässigbare rechtliche) Anforderungen, wie sie Cerf skizziert hat, also eine *digitale Architektur von Dauerhaftigkeit*, sondern auch
11. die Notwendigkeit der Schaffung einer *sozialen Struktur*, welche solch eine Architektur effektiv entwickeln, managen und weiter entwickeln kann. Offensichtlich sieht Kahn keine solche Struktur unter den gegenwärtigen, im weitesten Sinne »sozialen« Institutionen wie privater Firmen, Forschungsinstitutionen oder bspw. Bibliotheken. Ich komme darauf am Ende zurück.
12. Doch zuerst noch zu einem zweiten Ansatz, der zumindest die *sehr* langfristige Sicherung von Daten und ihrer IT-Umgebung verspricht. Sie stammt von Alan Kay, wie Cerf einer der Mitbegründer all dessen, was wir heute so an digitaler Technik benutzen – und einer ihrer schärfsten Kritiker. Falls Sie seine Vorträge und Interviews noch nicht kennen sollten, möchte ich Ihnen dringendst empfehlen, dies nachzuholen: Sie sind nicht nur sehr klug, voller Anregungen und oft überraschender Einsichten, sondern auch sehr unterhaltsam. Kay war einer der ersten Mitarbeiter des XEROS PARC, das Palo Alto Research Center der Kopiererfirma, in dem überwiegend junge Wissenschaftler um ab ca. 1968 machen konnten, was sie wollten. Die Ergebnisse dieses Teams beeinflussten die IT-Entwicklung wohl mehr als alles andere. Eine der frühesten Ideen Kays war das um 1970 konzipierte sog. Dynabook, das er hier als Modell präsentiert. Die Zeichnung stammt aus dem dazugehörigen Artikel. Nur zur Erinnerung: Damals gab es noch keinen *Personal Computer*, und einer der ersten,
13. der am PARC entwickelte »Alto«, war noch kein »Desktop«,
14. sondern selbst noch ein »Desk«
15. Der Alto war der erste Computer mit einer graphischen Benutzeroberfläche ... die eigentlich ursprünglich entwickelt worden war, um die Bedienung der XEROX-Kopierer zu vereinfachen. (Wenn Sie die Bedienoberflächen der aktuellsten Kopierer kennen, werden Sie das kaum glauben. . .) Das System arbeitete mit Fenstern und Objekten, die mit der Maus bewegt werden und in Ordern

zusammengefasst werden konnten: Alles so, wie wir es heute kennen . . . weil kurz nacheinander zwei junge Männer dieses System durch ihre Mitarbeiter nachbauen liessen, denn XEROX war an einer Lizenzierung nicht interessiert: Wer würde schon einen eigenen Computer haben und vielleicht sogar zuhause, also ausserhalb von Firma und Universität nutzen wollen? Ausserdem waren der Alto und Smalltalk ja eher nur die Spielzeuge dieser verrückten jungen Leute vom Palo Alto Research Center. Ach ja, bevor ich's vergesse: Die beiden jungen »Raubkopierer« der Ideen zur graphischen Benutzeroberfläche hiessen Steve Jobs und Bill Gates. . .

Nur eine Kleinigkeit: Diese anscheinend dreidimensional dargestellte Schlangenlinie, die sich da unten über Programmfenster und Desktop-Hintergrund schlängelt, ist keine nachträglich eingebaute Graphik oder ein Bildfehler, sondern eine der vielen Möglichkeiten zur graphischen Arbeit auf dem Alto und mit Smalltalk, die Jobs und Gates »vergaßen« oder nicht vermochten zu kopieren und die der Alto allen unseren modernen Systemen also immer noch voraus hat.

16. Dasselbe gilt übrigens für das »oN-Line System«, das Douglas Engelbart bereits 1968 in einer Präsentation vorstellte, die heute gern als »Mother of all Demos« bezeichnet wird: Der Erfinder der Maus arbeitet dabei live an einer SDS-940, die 40 Meilen entfernt an seinem Institut stand, wo zeitgleich weitere Mitarbeiter auf diesem Mainframe-Computer arbeiteten, der über einen Festplattenspeicher von sagenhaften 96 MB verfügte. Der Arbeitsspeicher lag je nach Ausbau bei ca. 1 MB. Auf diesem System konnte Engelbart nicht nur strukturierten Text auf dem Bildschirm schreiben, *bevor* er sich entschloss, diesen als Text- oder Tabellen-Dokument zu speichern, er konnte auch beliebige Stellen in diesen Dokumenten immer wieder anspringen und »verlinken«. Er konnte aber auch auf dem Bildschirm zeichnen oder mit einem Kollegen an einem anderen Standort gemeinsam an einem Dokument arbeiten. Das alles allerdings, ohne ein bestimmtes Programm für Textverarbeitung, Bildbearbeitung oder Tabellenkalkulation zu starten, dann ein Dokument zu erstellen oder zu öffnen und beides dann wieder zu schließen. 50 Jahre später sind die Computer und Betriebssystem, die ich kenne, nicht dazu in der Lage. . .
17. Nach diesem kurzen Ausflug in die Computergeschichte wieder zurück zum Thema, denn Alan Kay hat sich spätestens seit den 1970er ebenfalls Gedanken darüber gemacht, wie man eigentlich die Daten, die man mit diesen neuartigen Computern so erzeugt, dauerhaft speichern könnte. Das dargestellte Speicher-

medium ist jedoch keiner seiner Entwürfe, sondern 5000 Jahre älter — denn es geht ja immer noch um langfristige Datensicherung und deren Management. Die mesopotamischen Keilschrifttafeln, englisch: cuneiform, waren für Kay allerdings Anlass, das Problem langfristiger Speicherung und Verfügbarkeit mit der radikalen Frage anzugehen: Wie kann man sicherstellen, dass Daten auch noch von Archäologen in 5000 gefunden, identifiziert und wiederhergestellt werden können?

18. Das Ergebnis war dieses Paper, ebenfalls von 2015, über »digitale Keilschrifttafeln«: Ein Beispiel dafür sehen Sie rechts: Dabei handelt es sich, im Prototyp auch bereits hergestellte Scheiben etwa von der Grösse einer CD bzw. DVD aus superhartem Glas, auf deren einer Seite eine weitgehend selbsterklärende Beschreibung ist, die leicht entzifferbar sein soll. Sie soll den Archäologen der Zukunft die ersten Hinweise liefern, dass es sich um ein Artefakt handelt, für dessen Benutzung eine Art »Vergrösserungsgerät« notwendig ist, mit dem dann die weitere Beschreibung über das zu rekonstruierende Computersystem gelesen werden kann. Mit dieser Anleitung soll es »in einem Nachmittagsjob« möglich sein, auf einem Computer, der mindestens irgendwie mit Nullen und Einsen umgehen kann, die auf der Disk enthaltene virtuelle Umgebung samt der darin enthaltenen Software und der damit erzeugten Daten zu rekonstruieren. Natürlich befindet sich auf dieser Disk nicht zufällig eine Version von Smaltalk '72. — Aber ist das wirklich eine Lösung für die heraufziehenden Probleme? Ich fürchte, nein.
19. Ein bisschen erinnert diese Idee vielleicht an ein mit viel EU-Fördergeldern entwickeltes System, das angeblich ohne eine solche selbsterklärende Beschriftung auskommen soll. Es wurde in Grenoble entwickelt und besteht ebenfalls aus einer Disk aus Saphir-Glas, auf der sich jedoch in Mikroform nur Bilder von Textseiten und Abbildungen befinden. Das Symbolbild täuscht etwas, denn tatsächlich passen auf eine Disk wohl mehrere Hundert Buchseiten ... aber von einer Sicherung und Nutzbarmachung von Forschungsdaten oder umfangreichen Publikationen kann dabei wohl doch nicht ernsthaft die Rede sein: Oder wie speichert man Datenbankinhalte von hunderttausenden oder millionen Druckseiten Umfang auf solchen Disks?
20. Dieses System ist also vermutlich eher nur ein Kuriosum, aber die beiden kurz vorgestellten der Computerpioniere Cerf, Kahn und Kay haben m.E. auch gravierende Probleme:

- (a) Mir scheint, dass Cerfs *Digital Vellum* die herkömmlichen bzw. absehbaren Probleme mit der Stabilität und langfristigen Verwendbarkeit von IT-Systemen und ihrer Daten nur auf einer Meta-Ebene wiederholt. Aber vielleicht ist das gar nicht anders möglich.
- (b) Auch scheint die Vernetzung der Daten — ausgerechnet von den Erfindern des TCP/IP! — nicht wirklich berücksichtigt worden zu sein: Was passiert mit Bildern oder Dateiformaten, die von einem entfernten System geladen werden müssen? Nehmen wir an, die naheliegende Lösung wäre, alle in einem Dokument, einer Datenbank oder einer Webseite enthaltenen Elemente würden bei der Übertragung in ein *Digital Vellum* herunter geladen und mit gespeichert (ohne das System zu sprengen): Welches »Dokument« entspricht einer Facebook Timeline oder einer Twitter-Diskussion? In gewisser Weise lösen sich die Paradigmen, die unserer seit dem letzten Jahrhundert etablierten Metaphorik entsprechen, gerade vor unseren Augen auf: In diesem Schema bearbeiten wir Dokumente mit dafür geeigneten Werkzeugen namens »Programmen«, legen diese in hierarchisch gestaffelten Ordnern ab, die vielleicht Leitz-Ordnern und Regalen und Archivräumen entsprechen mögen; und alles findet auf *einem* Computer ab, der wie unser Haus mit der Welt über eine »aufgebohrte« Telefonleitung verbunden ist. An diesem Paradigmen orientiert sich m.E. das *Digital Vellum*. Aber schon die heute meistgenutzten IT-Geräte, Smartphones und Tablets und die darauf meistgenutzten »Programme« entsprechen diesem Paradigma längst nicht mehr oder nur noch in einer sehr beschränkt-spezifischen Sichtweise.
- (c) Das ungelöste Problem der weitestgehenden Verrechtlichung der Daten und ihrer Nutzung erscheint mir eigentlich als sogar unlösbar. Vielleicht wissen Sie, dass das ursprüngliche Ziel der jungen Google-Gründer die Schaffung einer digitalen Weltbibliothek war: Sie wollten »einfach« alle Bücher, Zeitschriften und Dokumente scannen und online verfügbar machen. Durch Texterkennung sollten sie indizierbar sein und DAFÜR wurde der Suchalgorithmus entwickelt, mit dem dann »nur noch« Webseiten indiziert werden konnten, weil die Heerscharen der Verlagsanwälte dieses Projekt »erfolgreich« verhinderten. Haben wir Grund zu der Annahme, dass ein zweiter Anlauf für das *Digital Vellum* erfolgreicher sein wird—jetzt, wo *Digital Restriction Management* viel weiter entwickelt ist als Ende der 1990er Jahre?
- (d) Und selbst wenn es uns gelänge, diese Probleme zu lösen: Wer soll das tun

und das *Digital Vellum* entwickeln, managen, weiter entwickeln und absichern? Genauer: Welche gesellschaftliche Institution im weitesten Sinne soll das tun? Eine Firma wie Google? Eine einzelne Nationalbibliothek? Die UNO? ... Gut, die NSA angeblich speichert alles, ohne zu fragen, aber das kann ja nicht das Fundament unseres Forschungsdatenmanagements sein.

- (e) Last but not least klingt es für mich noch nicht so, als wäre das nun schon mehrere Jahre in der Entwicklung befindliche *Digital Vellum* wirklich etwas, das sich problemlos auf Alltagscomputern nutzen liesse. Aber das dürfte nur eine Frage der Rechen- und Speicherkapazität sein, die sich ja rasant entwickeln.
- (f) Alan Kays *Digitale Keilschrifttafeln* sind von vornherein gar nicht als solch eine Arbeitsumgebung gedacht, muss man fairerweise eingestehen. Sie lösen höchstens das Problem der *sehr* langfristigen Datenaufbewahrung und bieten also nur
- (g) die reine ursprüngliche Datenmenge
- (h) Schon ihre Nutzbarkeit dürfte eingeschränkt sein: Wie kopiere ich ein Dokument von dieser Disk in mein eigenes System? Anders sind die Daten aber gar nicht veränderbar
- (i) Seit kurzem, also ca. 25 Jahre zu spät, gibt es ja nun auch in Deutschland (und der Schweiz) Versuche, ein Bewusstsein für die Problematik der langfristigen Forschungsdatensicherung und Institutionen zu schaffen, die Lösungen erarbeiten. Was spricht dagegen?
- (j) Nun, der Name. – Nein, das ist ein Scherz. Man kann den Initiatoren ja nicht ernsthaft die schöne Möglichkeit der deutschen Sprache zur Bildung langer Komposita vorwerfen. Aber das »national« gibt mir schon zu denken: Wessen Forschung und also Forschungsdaten sind denn heute noch, also: nur noch »national«? Ist die Befürchtung wirklich so abwegig, dass kurzzeitige politische Differenzen dazu führen könnten, in internationaler Kooperation gewonnene Forschungsdaten plötzlich nicht mehr an einem Ort oder im Rahmen eines »nationalen« Systems aufbewahren zu dürfen? (»Brexit, anyone?«)
- (k) Ich halte allerdings schon die zugrundeliegende Idee für problematisch, in Gremien, an denen möglichst alle interessierten Institutionen beteiligt sein sollen, Konzepte, Standards und Schlüssel für die Verteilung der

Mittel und IT-Kapazitäten zu erarbeiten: Mehr als 300 Hochschulen, 93 Leibniz-, 84 Max-Planck-, 80 Fraunhofer- und 18 Helmholtz-Institute. . .

21. Wie könnte ein anderer Lösungsvorschlag aussehen? Ich denke, um im IT-Sprech zu bleiben, ein vollständiger Reboot unserer IT-Infrastruktur, zumindest im Forschungsbereich ist notwendig. Dazu gehören:
- (a) Entwicklung eines vollständig freien und offenen Gesamtsystems aus Hardware, Betriebssystem, Anwendungssoftware und Datenformaten*
 - (b) Entwicklung und Pflege dieses Systems durch eine Institution mit «Ewigkeitsgarantie» vgl. Nationalbibliotheken, Nationalarchiven oder Nationalmuseen
 - (c) Bereitstellung des Gesamtsystems als portable Virtuelle Maschine für jedermann
 - (d) Änderungen am System sind nur nach Genehmigung durch Institution möglich
 - (e) damit soll eine größtmögliche langfristige Stabilität und Kompatibilität gesichert werden
 - (f) Die erstellten Daten sollten jederzeit in Repositorien übergeben werden können (auch als »Backup«), die von der Institution verwaltet werden, und
 - (g) weitestgehende Bereitstellung aller Daten zur freien Nachnutzung: denn «Information wants to be free!» (Richard M. Stallmann)
22. Wie würde, könnte oder müsste eine solche Infrastruktur aussehen: Sie sollte vor allem auf allen Ebenen möglichst einfache, sehr gut dokumentierte und also jederzeit reproduzierbare bzw. emulierbare Elemente verwenden, sowohl bzgl. Hard- als insbesondere auch Software. In gewisser Weise wäre sie ähnlich strukturiert wie das *Digital Vellum*, nur würde sie eben *nicht* versuchen, die Fehler, Probleme, verewigte Hotfixes usw. der Vergangenheit weiter mit zu schleppen und also zu reproduzieren: Auf einer einfachen Hardware-Basis (oder deren Virtualisierung) aber auch direkt auf allem, was mit Nullen und Einsen rechnen kann, könnte die virtuelle Maschine aufsetzen, in der dann Betriebssystem (soweit nötig) und Anwendungssoftware laufen und die alle Beschreibungen nötiger Datenformate etc. enthält. Mit Smalltalk, dessen Virtuelle Maschine wohl auf über 128 verschiedenen Hardware-Plattformen läuft, wäre ein gutes

Vorbild. Durch die Offenheit des Systems könnten Hersteller jederzeit entsprechende Systeme anbieten oder ihre angebotenen um die Möglichkeit erweitern, diese Umgebung darauf laufen zu lassen. An der Entwicklung könnte man gerade die IT-Institute der Universitäten und Forschungseinrichtungen beteiligen. Und der institutionelle Zwang, dieses System zu nutzen oder zumindest alle Forschungsdaten darin zu dokumentieren, könnte seine Durchsetzung beschleunigen. Die Vorteile würden die Nachteile überwiegen: eine Einschränkung der wissenschaftlichen Freiheit findet ja bereits jetzt statt, nur sitzen die Einschränkenden nicht bei der DFG oder beim SNF, sondern bspw. in Redmond und Cupertino.

Auf jeden Fall wird das ein langer und teurer Weg. Aber geht es kürzer, günstiger und trotzdem dauerhafter?

Der Vorschlag dürfte auf den ersten Blick als hypertroph und — hinsichtlich Umfang und Kosten — als geradezu grotesk erscheinen. Aber wäre die dauerhafte Entwicklung, Bereitstellung und Pflege eines *Digital Vellum* wirklich weniger aufwendig? Wäre es weniger kostspielig, bspw. von allen Daten eines jeden Forschungsprojekts regelmässig eine *digitale Keilschrifttafel* zu erzeugen, zu lagern und zwecks Wiederauffindbarkeit zu katalogisieren? Wäre eine verteilte Infrastruktur, die *alle* aktuell irgendwo verwendeten Daten, spezifische und auch Standard-Softwarelösungen bspw. für Experimente sowie notfalls Hardware-Emulationen an mehreren Standorten speichern und *wiederauffindbar* katalogisieren muss, wirklich so viel weniger aufwendig? Ich denke: nein! Selbst wenn eine der vorgestellten Lösungen kurzfristig günstiger sein sollte: Langfristig führt meiner Meinung nach kein Weg daran vorbei, eine solche Struktur zu entwickeln und mit dem jahrzehntelangen »Wildwuchs« aufzuräumen. Der war bereits kostenintensiv genug, nicht zuletzt, weil jedes Projekt sich seine eigene Lösung basteln musste. . .

Allein schon die einmalige Übertragung und Konvertierung der bisher erzeugten und für aufbewahrenswürdig angesehenen Daten in dieses System wird sich in kurzer Zeit amortisieren. Denn die bisherigen Lösungen sehen nur vor, diese Daten entweder einfach »nur so zu speichern« und erst bei Bedarf vielleicht später zu versuchen, sie auch wieder benutzbar zu machen — ein Aufwand, der erfahrungsgemäss dann doch gescheut wird —, oder aber die Daten eben regelmässig auf die neueste Systemumgebung zu konvertieren. Sie machen das regelmässig mit allen Ihren alten Word Perfect oder Word-Dateien. . . ? Aber auch mit Ihren Access-Datenbanken?? Oder, da wir ja über digitale Publikation reden: mit ihren 15-20 Jahre alten QuarkXPress-Dateien? Ich fürchte:

nein.

23. Aber was machen wir, bis wir eines der oben genannten Systeme oder vielleicht sogar das von mir hier skizzierte benutzen können? Die Forschung anhalten ist ja wohl keine Option. Aber wir sollten unbedingt einige grundlegende Anforderungen beachten, um die gegenwärtige Lage, unser aktuelles *digital dark age* nicht zu verschärfen und die Datenproduktion für das *information black hole* nicht zu vergrößern: Die Nutzung freier, gut dokumentierter und als Open Source verfügbarer Hard- und Software sowie Datenformate wäre ein guter Schritt in diese Richtung.

Aber bis zur Entwicklung einer Lösung sollten wir also wohl Vint Cerfs anfangs zitierten Rat befolgen: *If there are documents you really care about: Print them out?*

Womit wir auch endlich nach langen Umwegen wieder beim digitalen Publizieren und der langfristigen Sicherung von Forschungsdaten wären. Das Beispiel, das ich Ihnen abschliessend kurz vorstellen möchte, setzt vollständig auf freie Software: Betriebssystem Linux (aber die Lösung läuft auch auf Windows und MacOS, Programmiersprachen Python und etwa C, Web Application Server ZOPE samt Content Management System Plone und darin ein Modul namens ftw.book, prototypisch entwickelt von der Berner Firma *4teamwork* und im Google Summer of Code 2016 unter meiner Mitwirkung ein bisschen weiter entwickelt: Es erlaubt, wie jedes CMS, die Erfassung von HTML-Dokumenten, PDF-Dokumenten, Bildern u.a. webtauglichen Inhalten in strukturierter Form, also z.B. in Ordnern, und deren Ausgabe unter Verwendung von L^AT_EX als PDF oder im TEX-Format. Für die eigentliche Darstellung z.B. als druckfähiges Buch-PDF stehen dann unzählige TEX-Formatvorlagen zur Verfügung. D.h., man kann seine Daten im Web vollständig indizierbar (wenn man das möchte) oder auch »privat« bereitstellen, ständig aktualisieren und jederzeit in Sekunden daraus ein PDF erstellen, das guten typographischen Standards entspricht und bspw. über eine Print-on-Demand-Service als jederzeit aktualisierbares Buch hergestellt werden kann. Wenn ich gerade wieder erleben muss, wie der wissenschaftliche Publikationsalltag noch immer abläuft, scheint mir dieses System, dessen technische Voraussetzungen alle eigentlich schon seit ca. 25 Jahren vorliegen, meilenweit voraus zu sein.

Vielen Dank für Ihre Aufmerksamkeit und Geduld.