

## KATHEDRALE ODER BASAR?

### Überlegungen zu einer neuen IT-Infrastruktur (nicht nur) für die Digitale Kunstwissenschaft

**ZUSAMMENFASSUNG** Vor dem Hintergrund der bisher nicht gewährleisteten wirklich langfristigen Nachhaltigkeit heutiger IT-Infrastrukturen, wie sie in den Digital Humanities genutzt werden, vertritt der Artikel die Auffassung, dass eine schrittweise Anpassung heutiger Strukturen an zukünftige Entwicklungen über Jahrzehnte angesichts des unüberschaubaren „Wildwuchses“ in der Entwicklung von Hard- und Software kaum vorstellbar erscheint. Eine Lösung kann nur in einem bewussten Neustart in Form einer auf extrem langfristige Entwicklung ausgelegten, komplett freien Gesamtlösung aus freier Hard- und Software liegen. Solch ein epochales Projekt erfordert jedoch – wie ein Kathedralbau – langfristige Planung und ständige Kontrolle.

**SCHLAGWORTE** Digitalisierung, Forschungsdaten, Hardware-Entwicklung, Langzeitverfügbarkeit, Nachhaltigkeit, Software-Entwicklung

**ABSTRACT** Against the background of the hitherto not guaranteed truly long-term sustainability of today's IT infrastructures as used in the digital humanities, the article argues that a gradual adaptation of today's structures to future developments over decades seems hardly conceivable in view of the unmanageable uncontrolled growth in the development of hardware and software. A solution can only be found in a conscious restart in the form of a completely free overall solution of free hardware and software designed for extremely long-term development. Such an epochal project, however, requires – like a cathedral building – long-term planning and constant control.

**KEYWORDS** Digitization, hardware development, long-term availability, research data, software development, sustainability

## **EINLEITUNG**

Der Titel dieses Essays bezieht sich auf den inzwischen klassischen Text Eric S. Raymonds über die unterschiedlichen Entwicklungsmodelle für Closed und Open Source beziehungsweise Free Software, der 1997 erstmals in Würzburg auf dem 4. Internationalen Linux-Kongress vorgetragen wurde und später beim (kommerziellen) Verlag O'Reilly als frei verfügbares Buch im Open Access erschien.<sup>1</sup> Darin bezieht Raymond sehr deutlich Stellung für das im Bereich Open Source / Free Software etablierte Modell, das er mit einem Basar vergleicht: Viele Mitwirkende veröffentlichen ihre Beiträge so schnell und so oft wie möglich und natürlich in einer Form, in der sie für andere sofort nachnutz- und erweiterbar sind. Inwiefern dieses Modell wirklich mit dem metaphorischen Basar in Übereinstimmung zu bringen ist, bleibe einmal dahingestellt. Demgegenüber wird der Bau einer Kathedrale von Raymond als zentralistisches Entwicklungsmodell präsentiert, bei dem ein maßgeblicher Architekt und dessen engste Mitarbeiter nach einem fertigen, wohl sogar als nicht öffentlich einsehbar gedachten Plan an einem Großprojekt arbeiten, das sehr lange zu seiner Fertigstellung benötigt und anschließend nicht mehr oder kaum noch verändert oder erweitert werden kann. Auch hier könnte man – nicht nur als Architekturhistoriker – Zweifel anmelden, ob dieses Modell als Metapher für Großprojekte in der Entwicklung von Closed-Source-Software wirklich passend ist und mit der architekturhistorischen Realität übereinstimmt. Aufgrund von Raymonds Gegenüberstellung beider Modelle als sich weitgehend ausschließende Alternativen müsste der Titel aber eigentlich ein „or“ enthalten, nicht ein „and“ – und es sollte deutlich(er) werden, für welche Alternative der Autor plädiert. Da er dies unterlassen hat und ich seinen Titel nicht plagiiieren kann und will, bleibt kaum etwas anderes übrig, als hier also das „oder“ anstelle des „und“ zu verwenden. Dass dies sogar sinnvoll ist – insbesondere, wenn man die Alternative als Frage formuliert –, gerade weil ich beide Modelle für gerechtfertigt und beide zur Lösung des unten kurz skizzierten Problems für unverzichtbar halte, soll im Folgenden dargelegt werden. Ebenso möchte

---

1 Raymond, Eric S.: *The Cathedral and the Bazaar. Musings on Linux and Open Source by an Accidental Revolutionary*, Cambridge u. a. 1999, unter: [https://monoskop.org/images/e/e0/Raymond\\_Eric\\_S\\_The\\_Cathedral\\_and\\_the\\_Bazaar\\_rev\\_ed.pdf](https://monoskop.org/images/e/e0/Raymond_Eric_S_The_Cathedral_and_the_Bazaar_rev_ed.pdf) (abgerufen am 15.6.2020).

ich skizzieren, warum ich als Lösung des anderweitig bereits umfangreicher erläuterten Problems der meines Erachtens bisher nicht gegebenen, langfristigen Verfügbarkeit unserer digitalen Forschungsdaten, -ergebnisse und -software (im weitesten Sinne) eher das Modell des Kathedralbaus bevorzuge, welches aber unbedingt durch die Offenheit des metaphorischen Basars zu ergänzen wäre.<sup>2</sup>

## WARUM ÜBERHAUPT EINE NEUE IT-INFRASTRUKTUR?

„We are nonchalantly throwing all of our data into what could become an information black hole without realising it. We digitise things because we think we will preserve them, but what we don't understand is that unless we take other steps, those digital versions may not be any better, and may even be worse, than the artefacts we digitised. If there are photos you really care about, print them out.“<sup>3</sup> Mit diesen Worten charakterisiert Vinton „Vint“ Cerf, der gemeinsam mit Robert E. Kahn Anfang der 1970er Jahre das TCP/IP<sup>4</sup> entwickelte – die Gruppe von Protokollen für den Datentransfer, auf denen das Internet heute noch beruht –, die Situation unserer digitalen Daten: Sowohl die Datenformate, in denen sie gespeichert werden (Text-, Tabellen- und andere Dokumente, Fotografien und Bilder allgemein, Grafiken, 3D-Modelle, Animationen usw.), wie auch die Software, die zu ihrer Speicherung, Benutzung, gegebenenfalls Veränderung, Darstellung oder Konvertierung notwendig ist, haben – ebenso wie die Software, die das Funktionieren dieser Programme überhaupt erst ermöglicht Programmiersprachen und deren Compiler, also die sogenannte Middleware, vor allem aber Betriebssysteme) und die Hardware, auf der all das läuft und in Zukunft auch weiterhin laufen soll – aus dem Blickwinkel (nicht nur

2 Vgl. Kulawik, Bernd: Wie man das Verschwinden unserer Daten im „Digitalen Schwarzen Loch“ und somit ein „dunkles Informationszeitalter“ vermeiden könnte, in: Bienert, Andreas (Hg.): *Konferenzband EVA*, Berlin 2017, S. 220227, unter: <https://doi.org/10.11588/arthistoricum.443> (abgerufen am 15.6.2020); sowie: Kulawik, Bernd: Digitale Zwillinge sollten sich nicht zu sehr ähneln und 'getrennt wohnen', in: Bienert, Andreas (Hg.): *Konferenzband EVA*, Berlin 2018, S. 101–105, unter: <https://doi.org/10.11588/arthistoricum.442> (abgerufen am 15.6.2020).

3 Cerf, Vinton/ Sample, Ian (Hg.): (Interview) Google Boss warns of “forgotten century” with email and photos at risk, in: *The Guardian* [online], 13. Febr. 2015, unter: <https://www.theguardian.com/technology/2015/feb/13/google-boss-warns-forgotten-century-email-photos-vint-cerf> (abgerufen am 15.6.2020).

4 TCP/IP = Transmission Control Protocol / Internet Protocol. Zur Erläuterung von Funktion und Geschichte vgl.: [https://de.wikipedia.org/wiki/Transmission\\_Control\\_Protocol/Internet\\_Protocol](https://de.wikipedia.org/wiki/Transmission_Control_Protocol/Internet_Protocol) (abgerufen am 15.6.2020).

kunst-)historischer Zeiträume eine geradezu lächerliche Lebensdauer: Wie schnell sich Datenträger und Computer sowie deren Peripherie – Stichwort: Bildschirme – in den letzten 50 Jahren gewandelt haben und offenbar immer schneller wandeln, kann man fast täglich selbst beobachten. Für die Software gilt Ähnliches, auch wenn auf einigen museumsreifen Bankencomputern angeblich noch in den Programmiersprachen Cobol oder Pascal geschriebene Software aus den 1970er Jahren laufen soll, weil sich niemand traut, sie zu modernisieren – nach dem Motto: „Never change a running system!“<sup>5</sup>

Der Regelfall aber heute, den Cerf adressiert, dürfte – bezogen auf die Verwendung von Hard- und Software in Projekten der digitalen Kunstgeschichte – eher so aussehen: Die Hardware ist vermutlich fast so vielfältig, wie es der Markt hergibt, wobei meiner Erfahrung nach selbst teuerste Lösungen (HP Unix oder IBM-Mainframes mit Oracle-Datenbanken) in großen Institutionen angeschafft und betrieben, aber selten ausgelastet werden. Die meisten Institutionen werden jedoch aus Sparsamkeit PCs und Server „von der Stange“ betreiben, deren hardwareseitige Lebensdauer bei fünf bis maximal zehn Jahren liegen dürfte. Sollte es sich dabei nicht um bereits länger in Gebrauch befindliche Windows-Server handeln, wird darauf heute wohl eher eine der vielen gängigen Linux-Distributionen laufen, manche davon mit professionellem Support (zum Beispiel von Red Hat, SUSE oder einem kleineren Support-Drittanbieter), andere vielleicht eher nach dem Geschmack des jeweils zuständigen Administrators ausgesucht. Wer die Flame Wars zum Thema „Welche ist die beste Linux-Distribution?“ kennt, wird davon ausgehen dürfen, dass verschiedene Administratoren – gar über Generationengrenzen hinweg – eher nicht dieselbe Distribution bevorzugen: Hier sind also schon Umzugsprobleme in Form von Inkompatibilitäten vorprogrammiert, wie sie aber natürlich auch beim Upgrade von einer Windows-Version auf die nächste oder schon bei Updates innerhalb derselben Version gern auftreten. Damit einher geht aber auch bei diesen wie allen anderen Betriebssystemen und -generationen ein Austausch der Middleware, also zum Beispiel unterschiedlichste Versionen der mitgelieferten oder installierbaren Programmiersprachen, Gerätetreiber für die interne Hardware und die Peripherie, Office-Pakete oder Datenbank-Management-Systeme und so weiter und so fort. Bei aktuellen webbasierten

---

5 Dieses Beispiel ist übrigens schon ein hinreichendes Argument gegen den Glauben, es werde in Zukunft schon irgendwie möglich und irgendjemand bereit sein, heutige Daten regelmäßig auf dann aktuelle Systeme zu transferieren: Wenn Banken dazu nicht die Möglichkeit sehen und/oder die Finanzen bereitstellen, obwohl zumindest einige der Programmierer noch leben – wer dann?

Forschungsdatenbanken kommt dann gern noch die Abhängigkeit vom jeweils gerade aktuellen Browserstandard hinzu, die zu beachten wäre.

Wer sich einmal die Mühe machte, eine gängige datenbankgestützte, in der Regel ein freies CMS verwendende Lösung in einem typischen Projekt der Digital Humanities daraufhin zu untersuchen, von welchen Hard- sowie Software-Komponenten diese insgesamt überhaupt abhängig ist, wird schnell auf mehrere Hundert (Kernel-)Module, Dutzende Treiber, mehrere Skriptsprachen samt deren jeweiligen Parsern, die wiederum in verschiedenen Hochsprachen geschrieben sind, sowie häufig noch auf eine beliebige Zahl im Projekt selbst geschriebener Anpassungen in Form von Skripten kommen, die alle fortlaufend gepflegt, von in jedem Fall irgendwann zu entdeckenden Fehlern bereinigt und bei Inkompatibilitäten angepasst werden müssten: Das in den letzten Jahren erfolgte, sehr gut zu beobachtende „Einschläfern“ von 32-Bit- zugunsten von 64-Bit-Hard- und Software kann als gutes Beispiel dafür dienen, was in Zukunft immer wieder und vermutlich immer häufiger und auch immer schneller auf die Betreiber eines Projekts zukommt: Denn es ist ja nicht so, dass 32-Bit-Software auf 64-Bit-Hardware laufen würde, weil 32 doch irgendwie so etwas wie eine Untermenge von 64 ist!

Solange ein Projekt noch gefördert wird oder eine gewisse Anschlussfinanzierung gesichert ist, mag das noch – bis auf einigen personellen Aufwand – gut gehen. Aber kann jemand ernsthaft glauben, dass dies auch noch in 50 oder gar 100 Jahren für ein heute gestartetes und in absehbarer Zeit – meist nach drei bis zehn Jahren – beendetes Projekt funktioniert? Oder sind die von uns gesammelten Daten und daraus gewonnenen Erkenntnisse es vielleicht gar nicht wert, so lange aufgehoben zu werden? Zumindest diesen Eindruck könnte man angesichts der von Cerf kritisierten Unbekümmertheit im Umgang mit solchen Problemen bekommen. Dabei findet nicht einmal ein Umgang damit statt, sondern man schließt eigentlich eher ganz fest die Augen und hofft, dass alles schon irgendwie laufen wird. Deshalb empfinde ich es als Kunst- und Musikhistoriker geradezu beschämend, dass ausgerechnet beziehungsweise nicht einmal in den historischen Fächern an diese Fragen überhaupt nennenswerte Gedanken verschwendet zu werden scheinen.

Kurzum: In jedem Projekt haben wir es mit jeweils sehr idiosynkratischen Lösungen zu tun, die zudem oft genug nicht von hochqualifizierten Spezialisten geschrieben wurden und somit nicht gewissen Standards entsprechende Codequalität haben, weil solche Spezialisten im starren Tarifsystem öffentlicher Einrichtungen viel zu teuer wären. Stattdessen dürfte man zumeist einen regelrechten Zoo oder gar Dschungel vorfinden, für welchen

die Möglichkeit der langfristigen, dauerhaften Pflege – von Weiterentwicklung gar nicht zu reden – mit an Sicherheit grenzender Wahrscheinlichkeit ausgeschlossen werden kann.

Zwar gibt es Bemühungen, die Daten möglichst unabhängig – abstrahiert – von der spezifischen Hard- und Software zu speichern, sodass sie theoretisch problemlos beispielsweise als XML- oder JSON-Datei nach zuvor (!) genau zu definierenden Schemata zwischen verschiedenen Systemen ausgetauscht werden können – wenn diese sich an den gemeinsamen Standard halten. Aber meines Wissens gibt es bisher etwa noch keine zwei Implementierungen des CIDOC-CRM,<sup>6</sup> zwischen denen ein Datenaustausch erfolgreich versucht worden wäre. Auf jeden Fall benötigt man dafür in der Regel Parser, also eine weitere Komplexitätsschicht aus verschiedenen Skripten in einer (weiteren) Programmiersprache, die erst programmiert und dann gepflegt werden müsste.

Aber alle diese Probleme werden ja nun seit circa 2015 – und also mit „nur“ ungefähr 25-jähriger Verspätung<sup>7</sup> – endlich angegangen, indem Vorschläge für eine (oder mehrere?) Nationale<sup>8</sup> Forschungsdateninfrastruktur(en) von Gremien entworfen, entwickelt und gesammelt werden. Und irgendwann sollen dann sicher auch die Gewinnerprojekte dieses Arbeits- und Lebenszeit sowie jede Menge Gehirnschmalz verzehrenden Wettbewerbs finanziert und umgesetzt werden. Nach meinem Kenntnisstand dürfte es im Ergebnis des Prozesses eine oder mehrere landes- oder regionenweit verteilte Cluster-Systeme geben, auf denen Projekte dann ihre Daten und – das wäre zumindest zu verlangen – auch die zu deren Nutzung

---

6 CIDOC-CRM: CIDOC = Comité International pour la Documentation, CRM = Conceptual Reference Model, vgl.: [https://de.wikipedia.org/wiki/CIDOC\\_Conceptual\\_Reference\\_Model](https://de.wikipedia.org/wiki/CIDOC_Conceptual_Reference_Model) und <http://www.cidoc-crm.org> (beide abgerufen am 15.6.2020); es ist von der ISO als Standard beziehungsweise Norm 21127:2014 anerkannt.

7 Es ist mir rätselhaft, dass ausgerechnet die Digital Humanities sich nicht bereits zu *Beginn* der Digitalisierung – zumindest in ihrem Bereich und damit mit einiger zeitlicher Verzögerung zur technischen Entwicklung – kritisch fragten, ob und wie lange diese neuen Tools und die damit erarbeiteten Ergebnisse denn überhaupt „haltbar“ sein würden – wo doch historische und methodische Reflexion gerade zu jenen Grundlagen gehören, die immer als Unterscheidungskriterium genannt werden, wenn es wieder einmal gilt, die eigene Wissenschaftlichkeit der historischen, Geistes- und Sozialwissenschaften gegen – zumeist von außen finanziell begründete – Zweifel an der Sinnhaftigkeit ihres Tuns zu verteidigen.

8 Dabei dachte ich bisher, Forschung wäre schon längst viel zu international, als dass solche nationalen Alleingänge wirklich sinnvoll durchführbar wären, selbst wenn – oder gerade wenn – sie von einem wirtschaftlich, wissenschaftlich und politisch potenten Land vorgeschlagen werden, was von den kleineren Partnern beispielsweise in der EU eher als Druck interpretiert werden könnte.

und Interpretation notwendige Software hinterlegen können. Vermutlich wird es sich dabei um eine Art verteilte virtuelle Maschine handeln, in deren Containern dann quasi die gesamten Projekte – also Daten inklusive projektspezifischer Software – gespeichert werden. Interessant wird sein, ob und wie diese Daten nicht nur abrufbar, also „unverändert“ nutzbar sein werden, sondern zugleich auch gegen direkte Veränderung geschützt sein sollen und trotzdem zukünftig ergänzt und erweitert werden können. Im Prinzip müsste man für solche Ergänzungen – zum Beispiel wenn zu einem Projekt über barocke Deckenmalerei nach dessen Abschluss neue Quellen oder gar vielleicht durch Restaurierungen neu zutage getretene, bisher unbekannte Gemälde entdeckt werden – dann jeweils einen Klon der Originaldaten erstellen und zugleich garantieren können, dass alte und neue Daten jederzeit voneinander unterscheidbar sind und dabei auch synchron gehalten werden. Das scheint mir zumindest nicht trivial zu sein, wenn man voraussetzt, dass nicht alle Projekte hinsichtlich der Nutzungs- und Bedienungs- sowie Sicherheitskonzepte identisch sein dürften. Im Ergebnis erhielten wir dann in – optimistisch geschätzt – zehn bis 15 Jahren eine Gruppe von hoffentlich zueinander kompatiblen Lösungen – nennen wir sie einmal provisorisch „Digitale Archive“ –, in welchen eine Vielzahl von Projekten mit Daten und Software auf unterschiedlichen Host-Systemen gespeichert werden. Dann wäre sicherzustellen, dass nicht nur diese Host-Systeme – in sich vermutlich schon hochkomplex –, sondern auch die aufgenommenen Daten und ihre projektspezifische Software ständig gepflegt, gegebenenfalls – anhand guter Dokumentation – regelmäßig an Neuentwicklungen angepasst und auf neue Hard- und Software migriert werden können. Allein eine solche Migration kann dabei aber erfahrungsgemäß bereits für einzelne Projektdatenbanken sehr anspruchsvoll sein und mehrere Jahre dauern. Das heißt, eine große, generalstabsmäßig organisierte Gruppe von Informatikern und Fachwissenschaftlern, die verstehen, was in welchen Projekten warum und wie gesammelt wurde, würde „bis in alle Ewigkeit“ damit beschäftigt sein, diese ständige Migration vieler Tausend, wenn nicht sogar Millionen Elemente in Hunderten von abgeschlossenen Projekten anhand der notwendigen – heute meist jedoch gar nicht in ausreichender Qualität vorhandenen Dokumentation – zu bewältigen. Sollten diese Host- und die darin eingebetteten Projekt-Systeme der Digitalen Archive – wie zu fordern wäre – aus Open Source oder besser Freier Software bestehen, könnte man immerhin hoffen, dass nach Raymonds Basar-Modell eine Vielzahl beispielsweise durch die ehemaligen Projekt-Institutionen bezahlter oder zumindest unterstützter Freiwilliger diese ständigen Anpassungen und Fortentwicklungen in vielen kleinen Schritten vornehmen wird. Dies würde

aber wiederum auch von einem sozialen Faktor, einer sozialen Struktur, abhängen, wie es Cerfs Kollege Robert E. Kahn 2016 beschrieb:<sup>9</sup> Die heute üblichen Institutionen sind auch bei einer Existenz von einigen Jahrzehnten meines Erachtens viel zu kurzlebig, um die fortlaufende Bewahrung wichtiger digitaler Forschungsdaten und -ergebnisse in solchen skizzierten Host-Systemen langfristig sicherstellen zu können. Aber eigentlich kann ich mir aufgrund nun fast 35-jähriger Erfahrungen im Umgang mit Software leider nicht vorstellen, dass dieses Modell überhaupt wirklich dauerhaft funktionieren kann.

Dabei sind in diesem Szenario Probleme noch gar nicht berücksichtigt, die sich beispielsweise aus dem Ablauf von Lizenzen und Zertifikaten und gegebenenfalls deren notwendiger Ersetzung ergeben werden. Dieses letztere Problem wird meines Wissens auch in Vint Cerfs Vorschlag eines Digital Vellum genannten Systems noch gar nicht adressiert, obwohl er es in seinen Vorträgen gelegentlich erwähnt.<sup>10</sup> Das seit mehreren Jahren geplante und in Entwicklung befindliche, aber noch nicht fertige Digital Vellum entspricht einem solchen Host-System, das nicht nur sämtliche Daten und Software in sich aufnehmen, sondern auch die gegebenenfalls spezifische Hardware zu deren Betrieb simulieren können soll.

Und obwohl (nicht nur) in den Digital Humanities schon seit fast 20 Jahren beispielsweise die Forderung nach Offenheit im Allgemeinen – vorläufig vor allem in Form des Open Access und der Open Data – vertreten wird und sich (viel zu) langsam durchzusetzen scheint, werden beispielsweise Texte wie dieser üblicherweise immer noch mit einem bekannten kommerziellen Closed-Source-Programm erstellt – beziehungsweise deren Erstellung mit diesem Programm verlangt –, für welches man nicht erst für die ferne

---

9 Kahn, Robert E.: *Challenges and Opportunities for Digital Preservation* (Keynote auf der Konferenz IPRES 2016 in Bern), Abstract unter: [https://ead.nb.admin.ch/web/ipres2016/frontend/indexda54.html?page\\_id=1166](https://ead.nb.admin.ch/web/ipres2016/frontend/indexda54.html?page_id=1166). Dazu auch ein Bericht von Jürg Tschirren: *Der Vater des Internets*, auf der Seite des Schweizer Radio und Fernsehens (SRF), 5.10.2016, unter: <https://www.srf.ch/news/panorama/der-vater-des-internets> (beide abgerufen am 15.6.2020).

10 Im Prinzip müsste man von den Projekten, die in ein solches System dauerhaft aufgenommen werden sollen, verlangen, dass jede Abhängigkeit in Form von Lizenzen und Rechten nicht nur für alle enthaltenen Daten, sondern insbesondere auch für die Software auf langfristige Unabhängigkeit kontrolliert und gegebenenfalls sicher gestellt wird, dass problematische Fälle durch unproblematische ersetzt werden. Kann man ernsthaft erwarten, dass beispielsweise Kunsthistoriker alle Lizenzfragen und -probleme der im Hintergrund ihrer Projekte verwendeten Software abklären können? Z. B. die Frage, ob es bei der Kombination der unterschiedlichen Lizenzen verschiedener Betriebssysteme, Skriptsprachen und Anwendungssoftware Inkompatibilitäten geben könnte, die deren Verwendung und Übertragung auf andere (Betriebs-) Systeme vielleicht ausschließen?

Zukunft in 50 Jahren mit Sicherheit ausschließen kann, dass diese erstellten Texte dann noch im selben Layout oder überhaupt lesbar sein werden. Man sollte jedenfalls, so meine ich, lieber nicht davon ausgehen, dass eine einfache Übertragung des zugrunde liegenden XML-Textes in ein anderes Textverarbeitungssystem einfach so möglich wäre und der dabei zweifellos zu erwartende Verlust an Formatierungsvorgaben und -eigenheiten bedenkenlos und ohne Inhalts- beziehungsweise Bedeutungsverlust hingenommen werden kann. Und die Verwendung der Nachfolgeversion dieser Software wird man auch nicht einfach so umgehen können, wenn diese beim Neustart beispielsweise nach Übertragung auf eine neue Betriebssystemversion in 20 Jahren zuerst einmal verlangt, mit dem Server der Herstellerfirma verbunden zu werden, um die Gültigkeit des Installationsschlüssels prüfen zu können. Man sollte sich vielleicht auch besser nicht darauf verlassen, dass der Hersteller dann und in alle Ewigkeit bereit sein wird – so er noch existieren sollte –, für alte Softwareversionen einen Authentifizierungsserver samt (Gratis-) Lizenzschlüsseln zur Verfügung zu stellen. Schließlich will ein kommerzieller Software-Anbieter seine Nutzer dazu bewegen, immer wieder neu Geld für die aktuellste Variante seines Produkts zu zahlen.

Ein weiteres Problem dürfte im zu erwartenden Modell der Forschungsdateninfrastruktur in der Vernetzung der Daten bestehen. Es ist zwar noch nicht weit verbreitet, aber im System HTTP/WWW von Anfang an vorgesehen, dass alle Daten nicht immer wieder überall repliziert werden müssen, sondern verlinkt werden können. So können beispielsweise extern irgendwo im Web vorhandene Bilder jederzeit in eine Webseitendarstellung, etwa eines Datenbankinhalts *ad hoc* eingebunden und bei Aufruf der eigenen Seite aus dem Netz hinzugeladen werden. Das betrifft selbst schon Bilder, die im Rahmen eines Projekts auf einem eigenen, separaten Server gespeichert und von anderen aus via URL oder IP-Adresse abgerufen werden. Aber im Zuge einer – vielleicht bereits wieder abklingenden? – Bewegung weg von der eigenen Instituts- oder Projektwebseite hin zur aktiveren beziehungsweise schneller Aktivität vermittelnden Facebook-Präsenz darf man schon fragen, welcher zu speichernden (Gruppe von) Datei(en) denn beispielsweise eine Facebook-Timeline entspräche, die *ad hoc* und fortlaufend aus unterschiedlichsten Datenquellen zusammengesetzt und aktualisiert wird. Und diese Quellen selbst können wiederum durchaus auf weltweit verteilten Servern liegen. Ohne die Möglichkeit, die Facebook zugrunde liegende Software unabhängig nach- und unter fortlaufender Anpassung weiterzunutzen, wird man die dort abgelegten Daten schon mittelfristig als verloren ansehen müssen. Dies gilt aber nicht nur für die Facebook-Präsenz eines Projekts oder einer Institution, sondern eben grundsätzlich für jede Webseite, in

die externe Inhalte eingebunden werden. Das Problem der „toten Links“ ist ja kein neues ...

Alan Kay, einer der Väter der objektorientierten Programmierung sowie der graphischen Benutzeroberflächen und mit der technischen Entwicklung der letzten 50 Jahre mindestens so vertraut wie Cerf, hat gemeinsam mit Long Tien Nguyen einen Vorschlag<sup>11</sup> für die dauerhafte Speicherung von Daten und Software in einer Art eingefrorenem Zustand unterbreitet – also ohne die Möglichkeit direkter nachträglicher Veränderung: Seine *Digital Cuneiform Tablets* genannten Speichermedien – aktuell in der Form von CDs beziehungsweise DVDs ähnelnden Datenträgern gedacht – sollen über eine auf der Oberfläche lesbare Selbstbeschreibung verfügen, die es jedem interessierten Archäologen sogar der fernsten Zukunft noch erlauben soll, das zum Zugriff auf die Daten notwendige Hardware-System in kürzester Zeit zu rekonstruieren. Auch hier muss man voraussetzen, dass die so zu sichernden Projektdaten unverändert erhalten bleiben sollen und also – ohne Duplizierung auf einem neuen System und damit gegebenenfalls nötige Anpassungen – nicht veränderbar sein werden. Man mag einwenden, dass dies für unsere wichtigsten Quellen – Kunstwerke und andere Artefakte ebenso wie Primär- und Sekundärliteratur – ebenfalls gilt und wir damit bisher gut umgehen konnten: Aber wenn wir auf diesem Stand verharren würden, gäben wir meines Erachtens schon einen wesentlichen Vorteil computerbasierter Forschung überhaupt auf. Der andere wäre die Verarbeitung sehr großer Datenmengen; aber hier ließen sich natürlich auch wie bisher Daten und Ergebnisse zusammengefasst auf Papier „speichern“. Damit wären die Digital Humanities jedoch nur noch in dem Sinne digital, dass sie den Computer als bessere Schreibmaschine und für Berechnungsschritte verwendeten – was vielleicht gegenüber Dritten keine guten Argumente für die kostenintensive Finanzierung ihrer IT-Entwicklungen wären.

## EIN LÖSUNGSVORSCHLAG

Aus meiner Sicht müsste daher die Lösung der aufgezeigten Probleme einige grundsätzliche Änderungen im bisher üblichen und absehbaren Vorgehen beinhalten: Meiner Meinung nach ist angesichts des erwähnten

---

11 Nguyen, Long Tien; Kay, Alan: *The Cuneiform Tablets of 2015*, Viewpoints Research Institute, VPRI Technical Report TR-2015-004, Los Angeles 2015, unter: [http://www.vpri.org/pdf/tr2015004\\_cuneiform.pdf](http://www.vpri.org/pdf/tr2015004_cuneiform.pdf) (abgerufen am 15.6.2020).

„Wildwuchses“ im Bereich der Hard- und Software nichts weniger als ein grundlegender Neustart notwendig.

Wir befinden uns vermutlich in einer Situation, die der explosionsartigen Entwicklung des Buchdrucks in der ersten Hälfte des 19. Jahrhunderts vergleichbar ist: Sie machte es schlicht unmöglich, dass eine von einem Kleinstaat-Fürsten, einem Kloster, einer Stadt oder einer Privatperson betriebene Bibliothek alle halbwegs relevanten Neuerscheinungen gezielt sammeln konnte. Deshalb entstanden Staats- oder Nationalbibliotheken, deren Hauptzweck es war, alle in der Landessprache oder im jeweiligen Staat erscheinenden Drucke zu sammeln und zu erschließen. Hierzu wurden diese Bibliotheken – ähnlich wie Archive und später auch Museen – mit einer Art „Ewigkeitsgarantie“ durch eine ausreichende staatliche Finanzierung ausgestattet, um ihre dauerhafte Existenz sicherzustellen: Man stelle sich nur einmal vor, damals wäre die heute vorherrschende Förderung einzelner Projekte in kurzfristiger Finanzierung üblich gewesen.

Außerdem entwickelten diese Bibliotheken verbindliche Katalogisierungssysteme, um die standardisierte Erschließung des gedruckten Wissens zu ermöglichen. Auf die heutige Situation übertragen bedeutet dies meines Erachtens, dass wir eine mit „Ewigkeitsgarantie“ ausgestattete (inter)nationale Digitale Bibliothek benötigen, die für die Entwicklung und Pflege eines langfristig stabilen IT-Systems beziehungsweise einer Plattform verantwortlich wäre, das beziehungsweise die sowohl Hard- als auch Software umfasst. Der Planungs- und Entwicklungsprozess müsste vollständig offen und transparent sein, sodass Vorschläge eingebracht und Fehler oder Gefahren langfristig und früh erkannt werden könnten. Dabei sollten nicht nur heute bekannte, gravierende Mängel des bisherigen Hard- und Software-Designs ausgeschlossen beziehungsweise vermieden werden, die offene Hardware – bis hinunter zum Chip-Design – würde zugleich auch eine Monopolisierung verhindern. Die Software sollte – ähnlich den Vorschlägen Cerfs und Kays – aus einer virtuellen Maschine bestehen, die jedoch so klein und portierbar wie möglich zu sein hätte: Als Beispiel könnte die Programmiersprache Smalltalk samt ihrer Entwicklungsumgebung dienen, deren virtuelle Maschine auf über 100 Computerarchitekturen läuft. In, oder vielmehr über dieser virtuellen Maschine läge dann eine Schicht von möglichst wenigen, standardisierten und aufgrund bisheriger Erfahrungen sowie aktuellster Überlegungen zum Softwareentwurf optimierter Middleware aus Programmiersprache(n) und Grundfunktionen wie beispielsweise für die Hardware-Anbindung und die graphische Darstellung. Weitere Softwarekomponenten wie Office-Programme und Datenbank-Management-Systeme wären ebenfalls Teil des Gesamtsystems, dessen Nutzung dann

für staatlich finanzierte Projekte verbindlich sein müsste – ebenso wie die notwendige Rücksprache vor irgendwelchen projektspezifischen Änderungen. So könnte sichergestellt werden, dass die gesamten Daten eines Projekts und deren Strukturen jeweils wieder in das Hauptsystem übernommen werden könnten. Die Institution wäre dann in ständigem Austausch mit den Nutzern verpflichtet, eine möglichst vorsichtige, langfristig extrem stabile Weiterentwicklung zu garantieren, sodass praktisch für jede jeweils absehbare Zukunft eine Abwärtskompatibilität gewährleistet werden könnte.

Dieses Hauptsystem könnte durchaus ein verteilter Server-Cluster sein, der von der zu gründenden Institution und ihren Partnern weltweit verteilt – Stichwort: Redundanz – und parallelisiert, aber beispielsweise zur Sicherheit mittels unterschiedlicher, als Teile der Plattform aber definierter freier Hardware-Architekturen bereitgestellt werden könnte. Für Privatanwender und Firmen könnte dieses System ebenfalls – eventuell nicht kostenfrei – verfügbar sein, damit sie ihre Daten nach Abschluss eines Projektes dort dauerhaft ablegen und – vielleicht nach einer Sperrfrist – frei zur Verfügung stellen könnten.

Es liegt nahe, für dieses System eine Art objektorientierter Struktur anzulegen, die zugleich als Dateisystem wie auch als Meta-Datenbank und zur Ermöglichung einer transparenten Netzverfügbarkeit dienen würde, wodurch es möglich wäre, jedes einzelne Element – beispielsweise wie eine Datei in einem UNIX-System, wo eigentlich alles eine Datei ist beziehungsweise als solche behandelt wird – über einen URL-artigen Identifier anzusprechen. Damit könnte nicht nur gewährleistet werden, dass diese Datenobjekte, deren Größe geradezu beliebig granulierbar wäre, jederzeit über diesen Identifier zu finden wären, unabhängig davon, in welchem Hostsystem der Gesamtplattform seine virtuelle Maschine sich gerade befindet. Zugleich könnte so die Stabilität von Verlinkungen und damit die Verfügbarkeit externer Daten in Projekten sichergestellt werden. Angesichts der praktisch längst schon gegebenen Nicht-Druckbarkeit umfangreicher Forschungsergebnisse könnte dieses System als Webplattform zugleich zur längst fälligen freien Veröffentlichung von Forschungs(roh)daten und Ergebnissen dienen: Dabei wäre der Datenbankeintrag zum Format eines Bildes ebenso wie das Bild selbst, die wissenschaftlichen Kommentare dazu oder der Werkkatalog, in den es mit anderen Bildern aufgenommen wurde, genauso wie jede Forschungsabhandlung und die Kommentare der Peers der Scientific Community dazu adressier- und somit verlinkbar. Dieser objektorientierte Charakter würde es also nicht nur erlauben, wissenschaftliche Kommentare und Bewertungen zu jedem beliebigen Datenobjekt hinzuzufügen und selbst wiederum über eine jeweils eigene Adresse zu referenzieren,

sondern das System könnte auch von Grund auf so funktionieren, wie es Wikis eigentlich seit 1995 tun: Jeder Link erzeugt zugleich den Backlink zwischen verlinkendem und verlinktem Datenobjekt. Wissenschaftsbibliometrische Zählsysteme hätten so eine durchaus solidere Grundlage als heute – vorausgesetzt, man will an diesen überhaupt festhalten, wogegen es meines Erachtens mehr gute Gründe gibt als dafür. Vor allem aber wäre in so einem System eine größtmögliche Offenheit gewährleistet: Jede\*r Beitragende stünde mit seinem oder ihrem Namen für das Beigetragene, Clusterbildung in Form von Zitierseilschaften und Ähnlichem würde schnell erkennbar und selbst die so tatsächlich allen Peers offen stehende Bewertung von Forschungsanträgen und -ergebnissen wäre endlich anstelle der heutigen Beschränkung auf speziell ausgewählte, aber anonyme Personen – was meines Erachtens das Gegenteil des Peer-Begriffes ist – möglich.

Natürlich dürfte die Entwicklung eines solchen, auf bereits erprobten Konzepten beruhenden, aber im Detail vollständig neuen Systems mindestens einen dreistelligen Millionenbetrag erfordern – aber in Zeiten, in denen wir solche Beträge für die Anschaffung einzelner „Tötungsmaschinen“ ausgeben, die nicht einmal funktionieren, in denen mal eben größere Beträge für absehbar scheiternde Drohnen- oder Mautprojekte rausgeworfen werden, in denen zwei- und dreistellige Milliardenbeträge durch mutwillig nicht geschlossene oder sogar erst erzeugte Steueroasen und -schlupflöcher gestohlen werden können oder in die Rettung von Banken beziehungsweise die sie vertretenden Spekulanten mit ihren Casino-Wetten fließen – in solchen Zeiten kann niemand ernsthaft behaupten, dass ein dreistelliger Millionenbetrag für die dauerhafte Sicherung unserer gesamten wissenschaftlichen Arbeit nicht verfügbar wäre.

## FAZIT

Die Schlussfolgerung aus dem Gesagten sollte also weder als Parallele zu „[...] lasst uns einen Turm bauen, dessen Spitze bis an den Himmel reicht“ (1. Buch Mose, 11,4) verstanden werden, denn es geht gerade nicht darum, ein hypertrophes Vorhaben zu realisieren. Noch ist die Situation so hoffnungslos, dass man in Anlehnung an Luther mit Hoimar von Ditfurth sagen müsste: „So lasst uns denn ein Apfelbäumchen pflanzen“.<sup>12</sup> Stattdessen sollten wir uns die Zuversicht der Kathedralbauer früherer Jahrhunderte zum Vorbild nehmen und ein Projekt beginnen, dessen vorläufiger Abschluss

12 Ditfurth, Hoimar von: *So lasst uns denn ein Apfelbäumchen pflanzen. Es ist soweit*. Hamburg 1985.

vielleicht in vergleichsweise ferner Zukunft liegen und ein ständiges Ausbessern und Weiterbauen auf einer einmal als stabil erachteten Grundlage durch eine unbefristet existierende Bauhütte erfordern wird, dessen Ergebnis diesen Aufwand aber mehr als rechtfertigt. Denn meines Erachtens ist es besser, im Verlauf eines solchen Vorhabens eine „Kathedrale“ zu haben, zu der auch der „Basar“ fleißig in Form einzelner Module freier Software beigetragen haben wird und die – wie ihre Vorbilder in der Realität – auf ewig eine Baustelle sein wird, als auf einem stetig wachsenden Haufen aufwendigst hergestellter, aber zueinander vollkommen inkompatibler Bausteine zu sitzen, mit denen niemand in der ferneren oder auch nur der näheren Zukunft mehr etwas anzufangen weiß und die man deshalb, trotz allen Aufwands zu ihrer Herstellung, irgendwann einfach wegwerfen wird.

ORCID®

Bernd Kulawik  <https://orcid.org/0000-0002-2083-6118>