

1 Einleitung

Fast kaum ein Vorgang der heutigen Informationsgewinnung, -speicherung und -übermittlung kommt ohne elektronische Hilfsmittel aus. Die Informationstechnik, die so den Wissensumsatz beschleunigt und ganz neue Formen von Arbeit und Kooperation erlaubt, ist auch nach 50 Jahren noch wenig auf Nachhaltigkeit ausgerichtet: Sie bringt durch ihre rasante Weiterentwicklung ihre eigenen Produkte in Gefahr.

Diese Gefahr wird seit der Jahrtausendwende zunehmend wahrgenommen. Deshalb gibt es inzwischen eine ganze Reihe von Initiativen und Bestrebungen gerade von *Gedächtnisorganisationen*,¹ die sich zum Ziel gesetzt haben, den permanenten Zugriff auf die digitale Wissensbasis und das elektronische kulturelle Erbe sicherzustellen. Die meisten der in diesem Zusammenhang gewählten Strategien, *digitale Objekte* mit der technischen Entwicklung mitzuführen und für den permanenten Zugriff bereitzuhalten, setzen auf *Migration*. Für Datenträger, üblicherweise das simple Umkopieren der Bitströme auf das jeweils aktuelle Speichermedium, stellt sie das Mittel der Wahl dar. Jedoch muss auch der Inhalt der Bitströme für die jeweils aktuelle Form des Zugriffs angepasst werden. Diese Art der Migration impliziert die (wiederholte) Veränderung des digitalen Objekts selbst.

Die in Betracht gezogenen Strategien der funktionalen Langzeitarchivierung sollten deshalb Komponentenbündel enthalten, die jederzeit einen Zugriff auf das Objekt in seinem Originalzustand erlauben. Das kann bedeuten, dass diese Ergänzungen überhaupt erst eine spätere Darstellung der Objekte gewährleisten. Emulation hat eine Wiederherstellung einer Entstehungsumgebung zum Ziel, in der ein Objekt ursprünglich erzeugt wurde oder primär ablief. *Emulation* verhindert auf diese Weise, dass die *Primärobjekte* eines *digitalen Langzeitarchivs* verändert werden müssen.

Was generell für digitale Objekte gilt, ist speziell auch für die Strategie der Emulation richtig: Auf digitale Objekte kann nicht ohne technische Hilfsmittel zugegriffen werden. So wie für das Abspielen von Schallplatten der Plattenspieler aufgehoben werden muss, benötigt man ein Archiv an Werkzeugen und geeignete *Referenzumgebungen* als Bezugspunkte zur Wiedergabe digitaler Objekte. Während sich bereits die sinnvolle physische Erhaltung des technisch recht trivialen Gerätes Plattenspieler als nicht unproblematisch erweist, gilt dies erst recht für digitales Equipment. Bei geeigneter Wahl der Mittel kann jedoch der Fall einfacher liegen – man muss kein Hardwaremuseum aufbauen und über lange Zeit pflegen. Stattdessen kann dieses einem speziellen Softwarearchiv mit Hardwarenachbildungen, den *Emulatoren*, übertragen werden.

¹Spezielle, im Zusammenhang dieser Arbeit relevante Fachbegriffe, die im Glossar am Ende des Anhangs näher erläutert werden, sind kursiv gesetzt.

Damit ein solches Archiv seine Aufgaben erfüllen kann, sind einige Randbedingungen einzuhalten. Mit der Einstellung digitaler Objekte in ein Langzeitarchiv ergeben sich Anforderungen an deren Wiedergabe, denen seitens des Managements des Softwarearchivs Rechnung getragen werden muss. Ein solches Archiv muss geeignet bestückt, gepflegt und regelmäßig aktualisiert werden. Ein Teil der Archivobjekte, die Emulatoren, sind ebenso von Veraltung betroffen, wie alle anderen Primärobjekte auch. Sie spielen jedoch eine kritische Mittlerrolle als Schnittstelle zum jeweils gültigen Stand der Technik: Digitale Objekte aus einem Langzeitarchiv sind nur dann sinnvoll für die Anwender ihrer jeweiligen Zeit nutzbar, wenn sie in geeigneter Weise in der zu der Zeit verfügbaren technischen Umgebung zugreifbar sind.

1.1 Motivation

Das Gedächtnis der menschlichen Spezies speist sich aus Überlieferung. Die heutige Kultur ist nicht ohne ihre Wurzeln vorstellbar. Die Ursprünge reichen bis weit vor unsere Zeitrechnung zurück. So besitzt beispielsweise die europäische Kultur starke antike und christlich-religiöse Fundamente. Das heutige Wissen speist sich aus dem fortgesetzten Umsatz von Information. Dabei kommt es insbesondere darauf an, dieses Wissen nicht nur im relativ kleinen persönlichen Kreis weiterzugeben, sondern dieses auch über Zeitspannen zu transportieren, die über das eigene Leben und das mehrerer Generationen weit hinausgehen. Eine Verbreitung von Information über größere Räume hinweg konstituiert ein wesentliches Element der Vergrößerung der Wissensbasis.

Für die Verbreitung von Wissen über Raum und Zeit werden Informationsträger benötigt. Die Entwicklung der menschlichen Kultur und ihrer Befähigung zum Informationsumsatz sind eng mit diesen Trägern verbunden. Bereits die Bewahrung früher und mittelalterlicher Schriften gelang in vielen Fällen nicht ohne Auffrischung. Man kann immer noch die in Stein gehauenen Hieroglyphen auf der Stele des Hammurabi und als Wandinschriften in diversen Grabkammern oder Tempeln entziffern. So lassen sich heute noch über 4000 Jahre alte Pyramiden mit ihren Inschriften bewundern, die von der Macht damaliger Herrscher in Ägypten künden. Die notwendige Bedingung, dass man etwas über die "Heldentaten" eines Königs, das Alltagsleben seiner Untertanen erfährt, liegt darin begründet, dass es zum großen Teil gelungen ist, die Bedeutung der Zeichen der damaligen Schriftsprache zu entschlüsseln.

Die ursprünglichen Papyri² oder Tierhäute, auf denen Texte vor mehreren tausend Jahren niedergeschrieben wurden, sind heute oft nicht mehr existent oder kaum mehr entzifferbar. In diesen Fällen musste schon seit früher Zeit durch wiederholtes Umkopieren für den Fortbestand der Information gesorgt werden. Die Aufgabe des Erhaltens übernahmen beispielsweise im Europa des Mittelalters die Klöster. In diesen waren viele Mönche mit nichts Anderem beschäftigt, als Bücher "abzuschreiben" oder zu übersetzen, um so das gesammelte Wissen zu bewahren und mit den Mitteln der jeweiligen Epoche zugänglich zu halten. Nicht zu unrecht bemaß sich der Reichtum der Klöster am Umfang und Inhalt ihrer Bibliotheken.

Mit der Zunahme des Wissens reicht dessen lediglich mündliche Überlieferung nicht mehr aus. Sind am Prozess der Weiterverbreitung und Übermittlung nur wenige Individuen beteiligt, steigt zudem das Risiko der Verfälschung oder des Totalverlustes.

²Papyrus als Vorläufer des Papiers im Altertum, siehe hierzu auch den Überblick "Datenspeicher: Das Gedächtnis der Kultur" in [Spilker 2007].

Nicht jede kulturelle Äußerung hat sich für zukünftige Generationen erhalten lassen. Das hängt durchaus von den Möglichkeiten des jeweils erreichten Stands der technischen Entwicklung ab. Jede Art der Darstellung, jede Form der Information verlangt nach durchaus sehr unterschiedlichen Strategien der Konservierung.

Im Grunde hat sich an der Jahrhunderte alten Fragestellung nicht viel geändert: Auch in unserer Epoche geht es um die geeignete Bewahrung von und die langfristige Zugreifbarkeit auf Information, selbst wenn sich mit den Möglichkeiten moderner Technik die Art der Lösung dieses Problems verändert hat.

1.2 Einordnung und Fokus der Fragestellung

Die Zahl der Akteure auf dem Gebiet der Langzeitarchivierung digitaler Objekte ist im Laufe der letzten zehn Jahre deutlich angestiegen. Inzwischen bestehen etliche Initiativen zum Erhalt des elektronischen kulturellen Erbes. Die zunehmende Spezialisierung hat eine breite Palette an Fragestellungen³ identifiziert, die sich zur Einordnung dieser Arbeit wie folgt gliedern lassen:

- **Datenträger** - Für die mittel- und langfristige Datenablage zu archivierender Objekte müssen geeignete Methoden gefunden werden, die über einen ausreichend langen Zeitraum Informationen stabil erhalten können. In diesem Bereich existieren erfolgreiche Ansätze wie DIAS oder Kopal.⁴ Daher befasst sich diese Arbeit mit Datenträgern lediglich an Stellen, wo diese als Installationsmedien von *Software* eine Rolle spielen und geeignet archiviert oder virtualisiert werden sollten.
- **Datenformate** - Aufgrund ihrer unüberschaubaren Anzahl sind *Datenformate* zur längerfristigen Nutzbarkeit auf Standardisierung angewiesen. Denn mit der Festlegung im Sinne der Langzeitbewahrung geeigneter Datenformate für eine Vielzahl digitaler Objekte steigt die Chance, dass entsprechende Software zur Interpretation dieser Formate zu späteren Zeitpunkten noch erhältlich ist oder aufgrund des hohen Datenbestandes in den Archiven verfügbar gehalten wird. In diesem Bereich wurden Vorschläge wie das ODF⁵ gemacht und als Standards aufgenommen. Solche Bemühungen wirken sich erst langsam auf zukünftige Archivbestände aus und sind daher für die an dieser Stelle vorgenommenen Betrachtungen weniger relevant.
- **Metadaten** - Digitale Objekte erklären sich vielfach nicht aus sich heraus. Oft benötigt man eine Reihe von Zusatzinformationen. Es existieren inzwischen einige gute Vorschläge für Metadatenschemata. Sie sind unter dem Aspekt der geeigneten Erweiterbarkeit für die Objekte des später erörterten Softwarearchivs zu erwähnen.

³Siehe beispielsweise die in [Waugh u. a. 2000], S. 176 ff. in "Preservation Challenges" getroffene Unterteilung.

⁴DIAS-Projekt der Königlichen Bibliothek der Niederlande, siehe [van Diessen und Steenbakkers 2002], Kopal [kopal 2008]. Einen kompakten Überblick zu Forschungsansätzen und Initiativen bietet darüberhinaus Abschnitt 2.6 in Kapitel 2.

⁵Open Document Format – unbeschränkt offener Standard (ISO/IEC 26300:2006) für sogenannte Office-Dokumente, siehe http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office.

- Digitales Archiv - Es wird zu erörtern sein, wie ein Archiv digitaler Objekte aufgebaut sein sollte und welche zusätzlichen Inhalte, *Sekundärobjekte*, für zukünftige Zugriffe benötigt werden.
- Langzeitarchivierungsstrategien und ihre Erfolgsbedingungen - Hierfür ist zu klären, welche Maßnahmen und Arbeitsabläufe notwendig sind, um ein vertrauenswürdiges Archiv⁶ zu betreiben und was bei der Objektaufnahme und späteren Rückgewinnung beachtet werden sollte.
- Standards - Sie helfen bei der Strukturierung der Problemstellung und dem Austausch zwischen Institutionen. Sie spiegeln sich in Referenzmodellen und Metadatenschemata wieder.

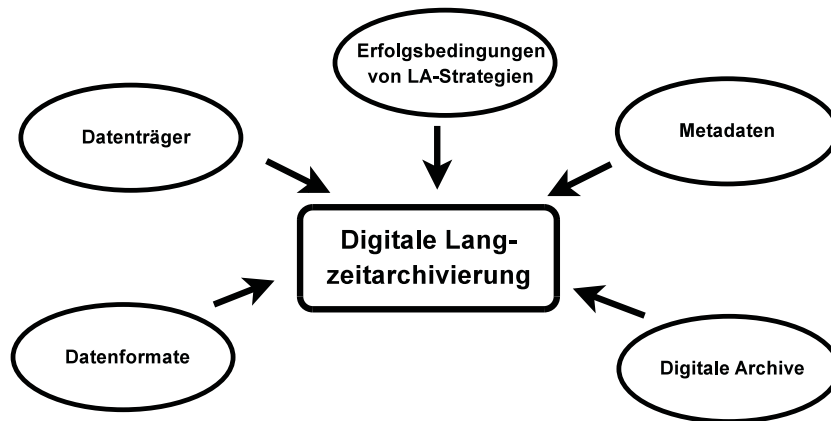


Abbildung 1.1: Das Problem der Langzeitarchivierung (LA) digitaler Objekte verschiedenen Typs lässt sich in eine Reihe von Teilfragestellungen ordnen.

Generell bedarf es einer Vervollständigung der bisherigen Ansätze, um es zu erlauben, digitale Objekte in ihrem Erstellungskontext zu behandeln. Dieser Kontext meint die Software- und Hardwarekonfiguration, in der das jeweilige Objekt erstellt oder zuletzt sinnvoll bearbeitet wurde. Diese Anforderungen erweitern das Archivkonzept und bringen eine Reihe neuer Aufgabenstellungen mit sich. Ausgehend von bereits bestehenden Überlegungen erfolgt im Rahmen dieser Arbeit eine Untersuchung und die Bereitstellung von Ablaufumgebungen. Für diese sind ebenso wie die ursprünglich zu erhaltenden Objekte Vorkehrungen zur langfristigen Fortschreibbarkeit zu treffen. Der Schwerpunkt dieser Arbeit liegt deshalb auf den Erfolgsbedingungen der Langzeitarchivierung einer Vielzahl digitaler Objekte und der Untersuchung, welche zusätzlichen Maßnahmen notwendig sind, um dieses zu erreichen. Hierzu werden Aufgaben und Anforderungen eines Archivs zur Emulation von *Rechnerplattformen* diskutiert.

Während die Anforderungen an Migrationsstrategien weitgehend bestimmt und die notwendigen Abläufe und Archivmanagementprozesse Untersuchungsgegenstand⁷ vieler Projekte und Initiativen sind, vermisst man diese Breite zur Emulation. In der bisherigen Forschung

⁶Vgl. hierzu die Nestor-Schrift [nestor 2006].

⁷Siehe 2.6 zu bisherigen Arbeiten und Untersuchungen - einen allgemeinen Überblick bringt Kapitel 2. Migration und Emulation werden in weiteren Abschnitten 3.4 und 3.5 im dritten Kapitel gewürdigt.

finden die speziellen Bedingungen dieses Zweigs der Langzeitarchivierungsstrategien noch wenig Aufmerksamkeit. Da sich die Emulation durchaus wesentlich in ihrer Handhabung, ihren Arbeitsabläufen und Erfolgsbedingungen von Migrationsprozessen unterscheidet, unternimmt diese Arbeit eine breiter angelegte Untersuchung zu Einsatz und Randbedingungen.

Die Verfolgung der Emulationsstrategie erfordert demnach eine ganzen Reihe von ergänzenden Überlegungen. Bezogen auf den Archivbetrieb, beispielsweise nach dem *OAIS-Modell*,⁸ unterscheiden sie sich in vielen Bereichen grundlegend vom Migrationsansatz:

- Emulation operiert nicht am Objekt selbst, sondern arbeitet an seiner Umgebung. Sie kann aus diesen Gründen eine sehr breite Palette digitaler Objekttypen abdecken. Dabei kann man nicht von der einen Strategie sprechen, sondern von verschiedenen Ansätzen mit derselben Zielrichtung.
- So ist bereits bei der Einstellung digitaler *Archivalien* in ein Langzeitarchiv festzustellen, welche Art von Ablaufumgebung benötigt wird. Diese Information sollte sich anhand der Objektmetadaten, beispielsweise unter Zuhilfenahme von Format-Registries ermitteln lassen. Dabei können verschiedene Wege zur Darstellung existieren, die allein aus Gründen der Zugriffssicherheit vorgehalten werden könnten.
- Ausgehend vom jeweiligen Objekt und dessen Anforderungen ergibt sich ein "View-Path", der sich aus Komponenten, wie Erstellungs-Applikation, benötigtes *Betriebssystem*, dafür geeigneter *Emulator* und geeignete Host-Plattform zusammensetzt. Wo Objekt, Applikation oder Betriebssystem gegebenermaßen weitgehend festliegen, sind Emulatoren anhand zu bestimmender Kriterien auszuwählen. Letztere bilden die Schnittstelle zum gewünschten Primärobjekt und dem jeweils aktuellen technologischen Stand, manifestiert durch die jeweilige Referenzumgebung.

Es reicht jedoch nicht, lediglich die ausgewählten Emulatoren zu archivieren. Diese müssen mindestens um Hilfsmittel für den Datentransport in und aus der Ablaufumgebung ergänzt werden. Neben dem eigentlichen, dem primären Objekt und seinen typischen Metadaten wird es deshalb erforderlich sein, ein sekundäres Softwarearchiv zu betreiben. Dieses beinhaltet die vom View-Path her vorgegebene und für die jeweils aktuelle *Nutzungsumgebung* geeignete Emulatoren. Zusätzlich zu den für den Datentransport verwendeten Hilfsprogrammen nimmt das Archiv alle für die geforderten View-Paths benötigte Software, Treiber und Patches auf. Dabei können im Archiv bereits vorbereitete Pakete für besonders häufig angeforderte View-Paths vorgehalten werden. Diese Pakete eignen sich wegen ihrer Kompaktheit zudem zum einfachen Austausch über verschiedene Gedächtnisorganisationen hinweg. Zudem lässt sich ein solches Archiv um Komponenten der automatischen Datenverteilung oder ein Rechtemanagement und Zugriffskontrollsystem ergänzen. Ein an die Archivobjekte selbst gebundenes Rechtemanagement ist kritisch zu hinterfragen.

In Abhängigkeit von der Archivbenutzergruppe, des Alters einer wiederherzustellenden Umgebung oder des gewählten Emulators sind ausreichende Betriebs- und Bedienungsinformationen vorzuhalten und gegebenenfalls laufend anzupassen. Jedoch haben Endanwender zu

⁸Open Archival Information System - Zentrales Referenzmodell für elektronische Langzeitarchive, siehe dazu [CCSDS 2002] und Abschnitt 3.1 im weiteren Teil der Arbeit.

Hause, Nationalarchive, Staats- und Universitätsbibliotheken oder technische Museen durchaus sehr unterschiedliches Betriebswissen und heterogene Anforderungen.

Hierzu greift die Arbeit auf bereits vorhandene Ansätze zurück, versucht diese zu strukturieren und weiterzuentwickeln. Das Ziel ist die Erstellung einer Handlungsanleitung im Sinne von Best-Practice-Beispielen. Zudem werden Elemente einer Softwarebibliothek für "digitale Archive" diskutiert. Deshalb beschäftigt sich die Arbeit außerdem mit:

- Konzepten digitaler Daten - Zu verschiedenen digitalen Objekten sind unterschiedliche Umgebungsbedingungen zu schaffen, um diese wieder zugreifbar zu machen.
- Softwarearchiven - Es soll ein Archiv für Computersoftware der unterschiedlichsten Rechnerarchitekturen diskutiert werden. In diesem erfolgt die Bewahrung von Programmen mit denen digitale Objekte ursprünglich erstellt wurden oder weitgehend optimal wiedergegeben werden können. Dieses erstreckt sich nicht nur auf die Applikation direkt, sondern muss die gesamte Nutzungsumgebung erfassen.
- Benutzerschnittstellen - Da ein Softwarearchiv für Emulation Endbenutzern die Möglichkeit geben könnte, ihre Daten selbst anzusehen oder sogar zu transformieren und zu bearbeiten, werden geeignete, leicht bedienbare Schnittstellen benötigt.
- Räumlicher Distribution und Redundanz von Archivdaten - Das benötigte Softwarearchiv umfasst viele Komponenten, die in mehreren Gedächtnisorganisationen erforderlich sind.

Diese Arbeit entstand im Rahmen der Arbeitsgruppe Langzeitarchivierung am Lehrstuhl für Kommunikationssysteme des Instituts für Information und des Rechenzentrums der Universität Freiburg. Die Arbeitsgruppe ist eingebettet in das PLANETS Project⁹ und kooperiert mit der Nestor Media AG.¹⁰ Sie arbeitet in den Teilprojekten "Preservation Action" insbesondere im Bereich "Strategy Development" und "Tools for Environments" mit und macht Vorschläge zur Einbettung von Emulation in das "Interoperability Framework"¹¹ Sie sieht sich als Ergänzung und im gewissen Sinne als Impulsgeberin zur Erarbeitung von Teilstrategien und Erfolgsbedingungen einer Langzeitarchivierung in erster Linie dynamischer Objekte.

Hierzu beschäftigt sich die vorliegende Arbeit mit den Randbedingungen der erfolgreichen Anwendung der Emulationsstrategie und beleuchtet dazu aus den eben genannten Punkten die Auswahlkriterien von Emulatoren unter dem Fokus der Langzeitarchivierbarkeit, die Bewertung von *Emulatoren* und *Virtualisierern* anhand dieser Kriterien, die Entwicklung und Auswahl von View-Paths und den hierfür notwendigen Sekundärobjekten. Hierzu wird ein Beispielarchiv diskutiert, welches die notwendigen Elemente enthält, um für eine Auswahl populärer Systeme Ablaufumgebungen für Endbenutzer anzubieten.¹² Die Überlegungen in

⁹EU-gefördertes Projekt zur Langzeitarchivierung, [PLANETS 2008]. Es wird unter dem "Information Society Technologies (IST) Programme" des Framework 6 anteilig finanziert (Project IST-033789).

¹⁰BMBF-geförderte Initiative Nestor [nestor 2008]. Siehe zudem Abschnitt 2.6 zu weiteren Aktivitäten auf dem Gebiet.

¹¹Vgl. hierzu Ausführungen zu PLANETS in Abschnitt 2.6.2.

¹²Aufgrund des schieren Umfangs vergangener und aktueller *Rechnerarchitekturen* und der darauf lauffähigen Software kann es diese Arbeit jedoch nicht leisten, ein vollständiges Archiv zu erstellen.