



Bachelorarbeit 1

**Lösungsstrategien zur digitalen
Langzeitarchivierung unter
Berücksichtigung des
OAIS-Modells**

**zur Erlangung des akademischen Grades
Bachelor of Science in Engineering**

Bachelorstudiengang IT Infrastruktur-Management

Eingereicht von: Christian Gossmann

Personenkennzeichen: 1410640014

Datum: 9. Februar 2017

Betreut von: DI Dr. Christian Büll

Vorwort

Eine Arbeit zur Darstellung der Probleme und Herausforderungen bezüglich Langzeitarchivierung ist in heutiger Zeit der Flüchtigkeit von digitalen Daten wichtiger denn je. Wenn damit ein Beitrag geleistet werden kann, auch nur einen einzigen Datenverlust zu vermeiden, dann hat sie sich schon bezahlt gemacht. Genau das war auch der Beweggrund, diese Arbeit zu schreiben. Dass Sie als Leser an digitalen Erhaltungsstrategien interessiert sind, haben Sie ebenfalls bereits dadurch zum Ausdruck gebracht, indem Sie gerade dieses Vorwort lesen. Damit ist der erste Schritt getan, Datenverlusten vorzubeugen und digitales Kulturgut über die Zeit zu retten.

Mich würde es sehr freuen, wenn Sie auch den Rest der Arbeit lesen würden und dort und da die eine oder andere Strategie zur Langzeitarchivierung beherrzigen. Genau um diese Strategien geht es auch, da sämtliche Kapitel allgemeingültige, zentrale Aussagen zur LZA beinhalten ohne zu sehr in technische Details abzuschweifen. Der Grund dafür liegt im Bestehen der grundlegenden Erhaltungsstrategien digitaler Daten. Dagegen lassen sich technische Details der sich ständig im Wandel der Zeit befindlichen Systeme in zehn bis 20 Sekunden Google-Recherche ermitteln, was nicht Sinn dieser Arbeit ist. Über Feedback zur Arbeit selbst würde ich mich natürlich sehr freuen. Erreichbar bin ich unter der E-Mail-Adresse: c@gossmann.at

Bedanken möchte ich mich für die Unterstützung durch die FH-Burgenland, allen voran bei meinem Betreuer DI Dr. Christian Büll sowie den vielen Personen, die sich bereits wissenschaftlich mit der Materie auseinandergesetzt haben. Auch meine Familie kam zu kurz, was hiermit wieder gutgemacht werden soll.

Christian Gossmann

Eisenstadt, 31. Jänner 2017

Inhaltsverzeichnis aktualisieren

Vorwort	ii
Kurzfassung.....	v
Abstract	vi
1 Einleitung.....	1
1.1 Problemstellung.....	1
1.2 Zielsetzung.....	3
2 Grundlagen.....	4
2.1 Bitstream / digitale Daten	4
2.2 Bitfäule.....	4
2.3 Dateisystem.....	5
2.4 Datei(format).....	6
2.5 Datenträger / Informationsträger.....	7
2.6 Digitales Objekt	8
2.7 Daten, Information, Wissen, Weisheit.....	8
2.8 Stand des Wissens	9
3 Methodik und persönliches Vorgehen	10
4 Digitalisierung als Konvergenz analoger Daten	11
5 Backup ist keine Archivierung – Unterschiede	14
5.1 Das 3-2-1-Backup-Schema.....	15
5.2 Das Generationen-Backup	16
6 Arten von Daten.....	18
6.1 Interpretationsregeln	18
6.2 Langzeitstabile Dateiformate.....	19
6.3 PDF/a (Texte, Rastergrafiken, Bilder).....	19
6.4 Software	21
6.5 Videoformate	21
6.6 Metainformationen	22
7 Datenträgerarchivierung – die Lösung?.....	25
7.1 Heute verfügbare langlebige digitale Datenträger	25
7.2 Die GlassMasterDisc	26
7.3 Die 5D Glass Disc	26
7.4 Lagerorte.....	27

7.5	Emulation	27
7.6	Migration	31
7.7	Spezialfälle	32
7.7.1	Der Refresh	32
7.7.2	Die Replikation	32
7.7.3	Repackaging	32
7.7.4	Transformation	32
8	Langzeitarchivierung nach OAIS	34
8.1	Das OAIS-Modell	36
8.1.1	Die Datenübernahme (Ingest).....	37
8.1.2	Die Datenaufbewahrung (Archival Storage).....	37
8.1.3	Datenmanagement	37
8.1.4	Systemverwaltung.....	38
8.1.5	Planung der Langzeitarchivierung (Preservation Planning)	38
8.1.6	Zugriff (Access).....	38
8.2	Kritik an OAIS	38
8.3	Die Besonderheit von OAIS.....	42
9	Ergebnis und Interpretation.....	44
10	Literatur.....	48
	Abbildungsverzeichnis	53
	Tabellenverzeichnis	54
	Abkürzungen.....	55
	Anhang.....	56
	Eidesstattliche Erklärung.....	57

Kurzfassung

Backups und Archivierung sind Ansätze zur Vorbeugung von Datenverlusten in einer immer mehr ausschließlich digitalisierten Welt. Leider reicht dies nicht, um Daten über längere Zeiträume auch sicher zu archivieren. Sogar die NASA verlor schon Daten der Saturnmission in den 70er-Jahren des vergangenen Jahrhunderts, da die alten Magnetbänder mangels geeigneter Lesegeräte nicht mehr nutzbar waren. Auch heutige Datenträger weisen nur eine begrenzte Haltbarkeit auf. Gesucht sind sowohl Strategien wie auch langfristig haltbare Trägermedien, um Langzeitarchivierung zu ermöglichen. Damit ist gemeint, Daten für künftige Generation im Sinne der menschlichen Kultur zu bewahren.

Leider ist es mit der Langzeitarchivierung allerdings auch noch nicht getan. Das beste Konzept nutzt nämlich nichts, wenn die archivierten Daten in hunderten von Jahren nicht mehr interpretiert werden können, da die zugrunde liegenden Dateiformate nicht mehr verstanden werden oder noch schlimmer, überhaupt kein Zugriff auf diese Daten erfolgen kann. Dies ist z.B. dann der Fall, wenn man nicht weiß, dass diese Archive - im Sinne physischer Lagerorte - auch existieren. Allerdings kommt wie beim vorhin erwähnte Fehlen der Magnetbandlesegeräte auch fehlende Hardware in Frage.

Schon heute ist es praktisch eine kleine Sensation, wenn man Laptops noch mit eingebauten Diskettenlaufwerken findet. Das Problem liegt hier im technologischen Fortschritt begründet, der es geradezu erforderlich macht, gesamte Umgebungen zu emulieren, in der alte Daten in „ihrer lauffähigen Software“ verarbeitet werden können. Alternativ ist man hier dazu gezwungen, ständig auf aktuelle Technologien zu migrieren. Betroffen davon sind sowohl Software selbst, Texte, Datenbanken, Spiele, Fotos und auch alte analoge Daten, die zunehmend digitalisiert werden, was sowohl schnelleres Auffinden und Verarbeiten wie auch geringeren Lagerplatzbedarf bedeutet. Hier wäre an Bibliotheken zu denken sowie dicke Bücher, bzw. Zeitungsarchive oder Videotheken mit alten VHS-Kassetten.

Lösungsansätze verspricht das OAIS-Modell, das eigentlich keinen Anspruch auf verbindliche Implementierungsstrategien erhebt, sich allerdings dennoch innerhalb weniger Jahre als De-facto-Standard etabliert hat. Wie in der Arbeit gezeigt wird, hat allerdings auch das OAIS-Modell seine Schwächen. Das OAIS-Modell behandelt digitale Daten und solche benötigen die erwähnten Langzeitdatenträger. Was Datenbanken anbelangt, basiert Archivierung selbst auf Datenbanken bezüglich der Referenzierung ihrer Objekte, bzw. stellen Datenbanken auch große Probleme hinsichtlich inkompatibler Datentypen dar. Man denke hier z.B. an die Unterschiede in Oracle- und MS SQL-Server-Datenbanken. Diese werden dann schlagend, wenn es an die LZA von Datenbanken geht, was natürlich auch andere Systeme als die vorhin genannt betrifft wie z.B. MySQL (Burleson Consulting, o.D.).

Abstract

This study examines backups and archiving as approaches in preventing data loss in a more and more digitalized world. The introduction outlines, that even the NASA already lost data from the Saturn mission in the 1970s, since the old magnetic tapes were no longer usable due to the lack of suitable reading devices. Today's data carriers also have limited lifetime. Strategies are needed as well as sustainable storage media for archiving purposes over a long period of time. This is to preserve data for future generations as part of human culture.

Unfortunately, long-term archiving alone is not enough. This paper therefore focuses on concepts of long time preservation of digital data. Best concepts are useless, if the archived data cannot be interpreted in hundreds of years from now, when the underlying file formats are no longer supported or even worse, no access to these data can be taken at all. This might be the case, if no one knows about these archives exist as physical storage places. However, like the mentioned magnetic tape readers belonging to the NASA, missing and compliant hardware is also possible.

The methodology section describes the systematic literature search as well as the personal approach for each chapter.

The findings indicate that at present time it is already sensational, if laptops are equipped with disc drives. The problem is technological advance, which makes it necessary to emulate entire environments, where old data can be processed in "their native executable software". Alternatively, you are forced to constantly migrate to current technologies. This includes software itself, texts, databases, games, photos and old analogue data, which become more and more digitalized for faster access and processing. Think of lower storage space requirements for libraries with their thick books or newspaper archives. Also videotheques with old VHS cassettes are to be mentioned.

Nevertheless, the OAIS model, which does not claim to be mandatory implemented, has established itself as the de facto standard within a few years. Furthermore, as it will be shown, the OAIS model also has its weaknesses. OAIS only focuses on digital data and these data require durable media for long time preservation.

Finally the conclusion gives advice in how to prevent data loss and how to organize digital data for archival purposes in the future.

1 Einleitung

1.1 Problemstellung

„Elektronische Medien sind nicht archivierbar“ (Stoll, 1998, S.263)

Mit Clifford Stolls Zitat könnte man die Arbeit eigentlich sofort beschließen. Stoll ist immerhin vom Fach, da er in den 70er-Jahren des vergangenen Jahrhunderts die Bandaufzeichnung von Daten der Pioneer-Sonde der NASA-Saturnmission überwachte. Wie jeder gute Systemadministrator speicherte er die Telemetrie zwar vierfach redundant, dennoch konnte einige Jahrzehnte später mangels noch funktionierender Lesegeräte nicht mehr darauf zugegriffen werden. Anzumerken ist dazu, dass Band- und sonstige Laufwerke dem Verfall von z.B. Gummiteilen für Antriebsriemen oder der Korrosion ihres Metalls unterliegen (Huth, 2009, S.8.27). Stolls Daten waren jedenfalls verloren (Stoll, 1998, S.263) und mit ihnen Mrd. an Dollar der Saturn-Mission, da sie ohne die Lesegeräte eben nicht mehr zu entziffern sind.

Dagegen sind historische Datenträger wie Papyrus, Steintafeln und Höhlenmalereien für das menschliche Auge auch ohne technische Hilfsmittel noch nach Jahrtausenden lesbar. Damit zeugen sie heute von den Pharaonen, den antiken Griechen, der Römer oder vom christlichen Evangelium. Ohne diese geschichtlichen Quellen wäre die Menschheit eine andere. Mehr noch: hätte man das Evangelium um das Wirken von Jesus Christus anstatt es niederzuschreiben schon damals ausschließlich digital auf CD-ROM produziert, wüsste heute wohl niemand, wer Jesus war – der Datenträger wäre infolge seiner Alterung und externer Einflüsse nicht mehr lesbar (Bárány, 2006, S.65).

Das Problem liegt nämlich neben den erforderlichen Lesegeräten auch in der im Laufe der Zeit auftretenden Verschlechterung der Lesbarkeit heutiger Datenträger, wie CD, DVD, Blu-Ray oder magnetisierten Trägerschichten auf nach wie vor in Verwendung befindlicher Bändern oder mechanischer Festplatten begründet. Entweder tritt hier sofortiger oder schleichender Datenverlust auf. Optische Datenträger verlieren dabei generell ihre Daten, wenn sie dem Sonnenlicht ausgesetzt sind und deren Hitzeeinwirkung trägt ihr Übriges dazu bei, um den Datenträger zudem physisch zu verformen, dass er sich nicht mehr in einem Laufwerk auslesen lässt (Bárány, 2006, S.75). Auch Magnetpartikel von Festplatten und Bändern heben sich im Laufe der Zeit gegenseitig ihre Magnetisierung auf und werden damit unbrauchbar (von Suchodoletz, 2009, S.27). Sogar moderne SSD-Festplatten unterliegen dem Elektrotod durch Spannungsspitzen von Blitzschlägen oder aus statischen Entladungen. Dagegen können Steine herumgeworfen oder Jahrtausende lang dem Sonnenlicht ausgesetzt sein, darauf eingemeißelte Daten bleiben erhalten. Auch auf zerknittertem Papyrus würden Daten lesbar bleiben.

Was fehlerhafte digitale Datenträger anbelangt, wäre konkret an Software zu denken, die nicht mehr korrekt arbeitet, wenn einzelne Maschineninstruktionen

falsch gelesen werden und somit fehlerhafte Programme die Folge sind. Selbst wenn das Programm noch laufen sollte, könnten falsche Werte produziert werden. Ebenso könnten falsch gelesene Werte schon als Eingangsoperanden dienen. Man denke z.B. an Bankkonten- und Steuerdaten oder behördliche Dokumente, die revisionssicher archiviert sein müssen (Ullrich [0], 2009, S.10.18). Sogar Leib und Leben können gefährdet sein, wenn medizinische Daten verarbeitet werden. Man stelle sich vor, ein Arzt erhält falsch eingelesene Blutgruppen oder ungültige Dosierungen von Medikamenten. Medizinische Daten, die heute mehr und mehr schon ab dem Kindesalter dokumentiert werden, sollen schließlich auch im hohen Alter korrekt zugänglich sein. Und da dies eben aus rein praktikablen Gründen fast nur mehr elektronisch anstatt in Aktenordnern erfolgt, trägt letztlich auch der Arzt dafür die Verantwortung (Spindler, Hillegeis, 2009, S.16.18).

Bereits hier wird deutlich, dass schon alleine im Laufe eines Menschenlebens mit Archivierungsfristen von bis jenseits von 100 Jahren hinsichtlich der Haltbarkeit digitaler Datenträger und -archive geplant werden muss. Historische Aufzeichnungen sollen natürlich noch länger der Nachwelt erhalten bleiben. Dem Datenverlust kann nämlich jeder zum Opfer fallen und dabei über mehrere Jahre und Jahrzehnte mit Liebe angesammelte Videos und Fotos verlieren, da diese heute ebenfalls fast nur mehr digital produziert werden (von Suchodoletz, 2009, S.14). Analoge Fotos würden zwar im Laufe der Zeit ausbleichen, aber zumindest noch vorhanden sein. Enkelkindern in einigen Jahrzehnten die Abenteuerexpedition des Großvaters zu veranschaulichen würde jedoch ohne entsprechende Langzeitarchivierungsstrategien digitaler Datenträger aber wohl nicht möglich sein (Neuroth, 2009, S.14).

Die Kernproblematik dabei ist nämlich, dass Lesegeräte und Verarbeitungsanlagen fähig sein müssen, gespeicherte Nullen und Einsen zu dekodieren, zu interpretieren und v.a. richtig als Information darzustellen. Selbst mit einer korrekt eingelesenen Bitfolge käme man in Jahrzehnten ohne Kenntnis des Dateiformats und ob es sich um Software, Werte oder Zeichen handelt, nicht zu diesem Ziel. Zwar kamen Historiker bei den ägyptischen Hieroglyphen und Höhlenmalereien ebenfalls nicht um eine Übersetzung herum, im Gegensatz dazu muss allerdings festgestellt werden, dass heutige proprietäre Dateiformate nicht einmal einzelne Zeichen offenbaren. Der Grund hierfür liegt vielfach in kommerziellen Interessen der Hersteller, Kunden an ihre Software zu binden, wo erst Dekodierung mittels entsprechender Lesesoftware ermöglicht, an Informationen zu gelangen (von Suchodoletz, 2009, S.21f.). Von gespeicherten Bildern, die ebenfalls in Formen von Nullen und Einsen kodiert sind - im Gegensatz zur sofort betrachtbaren Höhlenmalerei oder analoger Fotos - ist dabei noch gar nicht die Rede. Lesegerät, Software, Dateiformat und Ausgabegerät müssen nämlich perfekt aufeinander abgestimmt sein und nur ein einziges Problem in dieser Kette hieße: kein Zugang zu den Daten und keine daraus ableitbaren Informationen.

Dass digitale Speicherung von Daten somit generell unzuverlässig ist, haben auch die Hersteller entsprechender Systeme erkannt. Je nach Medium und Hardware wurden daher Fehlerkorrekturalgorithmen mittels verteilt über den Datenträger gespeicherter Prüfsummen implementiert, die bei Verschlechterung der Lesbarkeit eine Rekonstruktion nicht lesbarer Bits aus dem fehlerhaften Bereich ermöglichen, d.h. ähnlich wie bei RAID-Systemen. Als Klassiker wäre ein Kratzer auf einer CD zu nennen (Bárány, 2006, S.73). Ist allerdings der Toleranzwert überschritten, wenn der Kratzer zu breit, zu lang oder zu tief ist, sind die Daten verloren (Tischer, 1998, S.295).

1.2 Zielsetzung

Ausgehend von den genannten Problemen und der totalen Durchdringung quasi aller Lebensbereiche durch digitale Speicherung von Daten (von Suchodolez, 2009, S.14) stellt sich die zentrale Forschungsfrage, wie man denn diese langfristig, d.h. über Jahrhunderte hinweg sicher archivieren kann und welche Dateiformate sich dafür anbieten. Damit einhergehend sind Probleme der Integrität, Authentizität, Vertraulichkeit und Verfügbarkeit verbunden (Dobratz, Schoger, 2009, S.5.4.).

Vor dem Hintergrund des OAIS-Modells sowie entsprechender Normen und Standards (ISO 14721, ISO 16363, DIN 31644) soll gezeigt werden, wie Texte, Bilder, Musik, Filme, Software, bzw. Computerspiele langzeitstabil archiviert werden. Nach Beschreibung des aktuellen Forschungsstands soll eine Lösungsstrategie für zukünftige und zuverlässige Speicherung von Daten sowie zur sicheren Lagerung der Datenträger gefunden werden. Dass dieses Unterfangen nur gemeinsames Ziel mehrerer Akteure sein kann, darüber ist man sich schon heute einig. Einzelne Institutionen, bzw. Bibliotheken sind der gewaltigen Herausforderung einer LZA nämlich alleine nicht gewachsen, weshalb es nationaler Kooperations- und Austauschmodelle bedarf (Jehn, Schrimpf, 2009, S.2.6.). OAIS ist ein wichtiger Ansatz dazu. Ziel ist ferner, Online-Quellen auf langlebigen Trägermedien sowie die Arbeit selbst in einem langzeitstabilen Format zu speichern. Welches dies sein wird, soll Teil des Ergebnisses sein. Digitalversionen ermöglichen zudem schnellere und gezieltere Informationssuche als Papierexemplare (Bárány, 2006, S.29).

Dagegen werden rechtliche Problematiken der LZA hier nicht behandelt. Dabei wäre z.B. an Löschverpflichtungen beim Datenschutz zu denken, denen man nicht entsprechen könnte, wenn auf nur einmal beschreibbaren Datenträgern für hunderte und tausende Jahre archiviert werden würde. Auch beim Urheberrecht und dem Digital Rights Management gibt es erhebliche Probleme (Egger, 2005, S.97ff.) in Verbindung mit LZA, die allenfalls nur kurz erörtert werden können. Ferner ist es nicht Ziel, veraltete Datenträger aufzulisten und seitenweise Tabellen mit deren Spezifikationen zu füllen. Stattdessen soll der Schwerpunkt auf langzeitstabilen Formaten und künftigen Datenträgern sowie -systemen liegen.

2 Grundlagen

Um digitale LZA und ihre Probleme zu verstehen sind folgende Begriffe von zentraler Bedeutung:

2.1 Bitstream / digitale Daten

Digitale Daten bedeuten Vorliegen in Form von Nullen und Einsen. Dies entspricht den technischen Gegebenheiten eines elektronischen Verarbeitungsendgerätes, egal ob Computer, CD-, bzw. DVD-Player, Smartphone, etc... Der Bitstream ist dabei nur der Sammelausdruck für die Abfolge dieser in den beiden – daher binären – Zuständen kodierten Daten. Genau diesen Bitstream samt zugehöriger Dekodierungs- und Darstellungsmöglichkeiten auch Jahrhunderte lang zu erhalten, ist das Um- und auf der digitalen Langzeitarchivierung. Ohne den Bitstream sind nämlich sämtliche weiterführenden Archivierungsstrategien sinnlos (Ullrich [0], 2009, S.8.3). Daten können dabei alles repräsentieren:

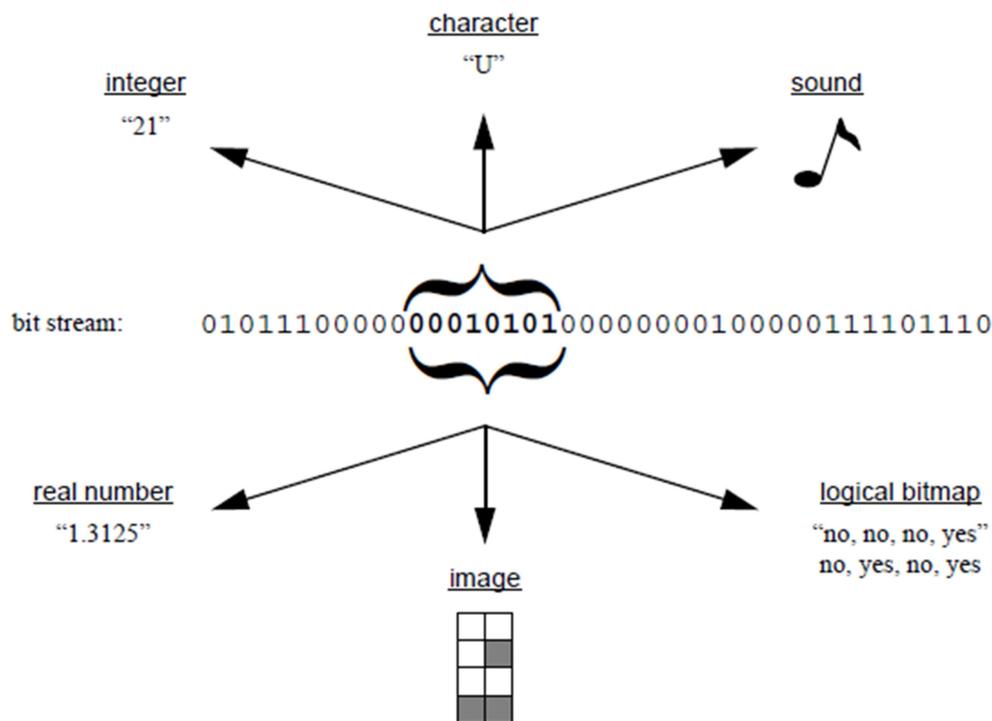


Abbildung 1: Was Bits aus Nullen und Einsen darstellen können (Rothenberg [0], 1999)

2.2 Bitfäule

Grund für den Erhalt des vorhin genannten Bitstreams ist die Bitfäule. Darunter versteht man sog. „umgefallene Bits“ in einem Bitstream. Diese mögen aus verschiedensten Gründen herrühren. Einerseits mag der Bitstream schon fehlerhaft gespeichert sein, andererseits kann er auch fehlerhaft eingelesen werden. Dies ist v.a. dann der Fall, wenn physische Beschädigungen eines

Datenträgers (siehe unten) auch die Fehlerkorrektur beeinflussen, die dann entsprechend fehlerhafte Bitstreams liefert. Trotz sog. Error Correcting Codes sind Fehler hier nämlich möglich (Windeck, 2013, S.177). Gefährlich wird dies v.a. bei Software, wo ein fehlerhaftes Bit bereits einen anderen Maschinenbefehl für den ausführenden Prozessor des Computers symbolisiert. Das Ergebnis reicht dann von Abstürzen bis hin zu Fehlberechnungen, sollte z.B. statt eines ADD-Befehls für Summe, ein SUB-Befehl zur Subtraktion dekodiert werden. Ähnliches gilt für Texte, wo falsche Zeichen zu Rechtschreibfehlern oder in Rechnungen zu falschen Ergebnissen führen, bzw. bei den einleitend erwähnten Medikamentendosierungen Gefahr für Leib und Leben droht. Zum Spaß würde wohl Österreichs Gesundheitsministerin Oberhauser nicht sagen: „Lieber ein bisschen länger testen, damit die Software funktioniert und die Datensicherheit gegeben ist, als sich an einen fixen Zeitpunkt zu halten“ (nach ORF [0], 2016).

Nachfolgend findet sich ein Beispiel digitaler Bitfäule:

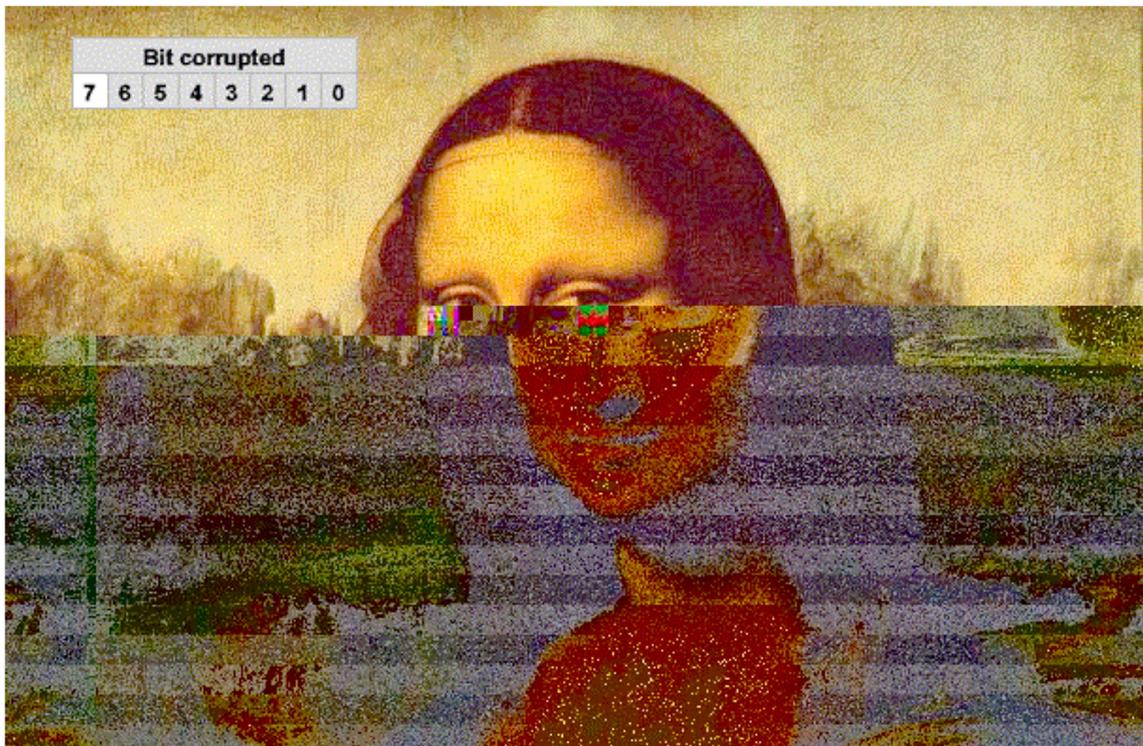


Abbildung 2: Digitale Bitfäule und ihr Ergebnis (Brandt, 2009)

2.3 Dateisystem

Dateisysteme stellen die Grundlage für die Speicherung von Bitstreams in Form von Dateien dar. Sie dienen der Organisation und der Wiederauffindbarkeit, aber auch der Trennung voneinander auf Datenträgern. Ferner lassen sich damit Zugriffsberechtigungen erteilen und verweigern (je nach Dateisystem). Im Anhang findet sich eine Darstellung mit den derzeit gängigen Dateisystemen. Ohne Dateisystem würde Chaos vorherrschen, denke man z.B. an den Verlust eines Inhaltsverzeichnisses und von Kapitelüberschriften in

einem Fließtext. Die Rekonstruktion der Dateien ist dabei zwar möglich, aber sehr mühsam und basiert darauf, dass jede Datei auch über Header-Daten verfügt. Fehlen solche Header, so „kleben“ die Bitstreams der Dateien aneinander und eine Rekonstruktion der Dateien ist kaum mehr möglich. Schwierigkeiten bereiten hier zudem fragmentierte Speicherung, d.h. dass einzelne Teile der Dateien in unterschiedlichen Sektoren von beispielsweise Festplatten gespeichert werden. Fragmentierung tritt dabei durch Beschreiben und Löschen von Dateien auf, wo deren Platz lediglich als neu zum Beschreiben markiert wird. Passen dann neue Dateien nicht zur Gänze in diesen Bereich, wird in freien Bereichen mit der Speicherung fortgefahren und es werden nur die entsprechenden Sektoren verlinkt. Noch gefährlicher sind SSD-Festplatten, die durch Algorithmen versuchen die Schreibzyklen auf ihre Flash-Zellen gleichermaßen zu verteilen. Wo nun eine Datei physisch gespeichert ist, „weiß“ nur die SSD selbst und kein Dateisystem. Datenrekonstruktion auf SSDs ist daher fast ein Ding der Unmöglichkeit. Auch Magnetbänder weisen kein explizites Dateisystem auf, sondern sequentielle Bitströme, ebenso Festplatten und Flash-Speicher. Das Dateisystem selbst setzt erst auf der logischen Ebene des jeweiligen Betriebssystems auf, das den Datenträger formatiert hat (von Suchodoletz, 2009, S.23). Mikrofilme dagegen haben ebenfalls kein Dateisystem im klassischen Sinn. Dennoch tendiert die LZA dazu, sich von Dateisystemen zu abstrahieren und Objekte wie bei OAIS in den Vordergrund zu stellen. Bárányi hat dem jedoch entgegen, dass, zumindest bei Open-Source-Dateisystemen dieses Problem weniger bedeutsam ist, da ja eben das Betriebssystem und damit auch die Lesemöglichkeiten der jeweiligen Dateisysteme hinreichend dokumentiert sind (2006, S.52).

2.4 Datei(format)

Was Dateiformate anbelangt, so ist auch hier Open Source zu bevorzugen, allerdings stellt auch der Grad der Verbreitung hinsichtlich der LZA und der damit verbundenen Auswahlmöglichkeit für zukünftigen Erhalt eine nicht zu vernachlässigende Bedeutung dar. Begründet wird dies damit, dass für sich einmal etablierte Dateiformate eigentlich keine Gründe bestehen, deren Unterstützung aus entsprechenden Verarbeitungsanlagen und Software wieder zu entfernen (Bárányi, 2006, S.53). Während dabei vorhin das Dateisystem als Organisationsstruktur für die Aufnahme der Bitstreams einer Datei selbst erläutert wurde, repräsentieren Dateien diesen Bitstream selbst. Dateien werden dabei anhand einer logischen Struktur innerhalb des Dateisystems abgelegt. An Dateien und Formaten gibt es unzählige Varianten wie z.B. für Textdateien, die einer entsprechenden Kodierung folgen, nämlich z.B. ASCII, ANSI, Unicode (Huth, 2009, S.17.5), Programmdateien, die unter dem Betriebssystem Windows die Kennung „exe“ aufweisen und mit der Kennung „MZ“ eingeleitet werden während grafische JPEG-Dateien hier an Byte-Position 8 die „JFIF“-Kennung aufweisen, wogegen komprimierte ZIP-Dateien mit „PK“ beginnen, usw... Verlassen darf man sich allerdings nicht auf diese Strukturen.

Es könnte sich auch um proprietäre Datendateien handeln, die zufällig an diesen Stellen diese Kennungen aufweisen, allerdings völlig anders zu interpretieren wären. Leider geben kommerzielle Hersteller, die meist auch Patente an eigenen Entwicklungen halten, die Spezifikationen „ihrer“ Dateiformate entweder gar nicht oder nur unzureichend bekannt - wenn, dann gerade so weit, damit das Format vermarktbar ist, indem man Schnittstellen für dessen Programmierung zur Verfügung stellt (von Suchodoletz, 2009, S. 23).

2.5 Datenträger / Informationsträger

Ein Datenträger, auch Informationsträger genannt, ist die physische Speichereinheit für digitale Informationen. Grund dafür ist, dass digitale Daten im Hauptspeicher eines Computersystems beim Ausschalten verloren gehen, also flüchtig sind. Um diese daher dauerhaft zu speichern, braucht man Datenträger. Als gängige Datenträger kommen derzeit magnetische (Band, Disketten, Festplatten), optische (CD-ROM, DVD, Blu Ray) oder Flash-Datenträger (USB-Stick, SSD-Festplatte) in Frage. Problematisch dabei ist deren unterschiedliche Lebensdauer, die bei häufigem Gebrauch auch noch abnimmt. Folgende Werte verstehen sich dabei als Richtwerte. In der Praxis ist jedenfalls mit weit darunter zu erwartender Haltbarkeit zu rechnen, mit Ausnahme des Mikrofilms.

Datenträger	Haltbarkeit in Jahren
Filme aus Zelluloid	100
Optische beschreibbare Speicher (CD-ROM, DVD, DVD-RAM, Blu Ray)	max. 50 unter Laborbedingungen
Gepresste optische Speicher (CD-ROM, DVD, DVD-RAM)	max. 80-100 unter Laborbedingungen
Festplatten im täglichen Einsatz	5
Festplatten als Backup	10
Flash-Speicher	10
Holographische Speicher	50
Mikrofilm	500

Tabelle 1: Haltbarkeit von Datenträgern (Rieder, 2013)

2.6 Digitales Objekt

Ein digitales Objekt stellt im Unterschied zu einem Bitstream, bzw. einer Datei eine genormte Struktur dar, die aus mehreren Bitstreams, bzw. Dateien bestehen kann. Vereinfacht ausgedrückt handelt es sich dabei um eine Art Paket. Vorstellen kann man sich dies als Container für mehrere Dateien, die zu einer ganzen, allerdings speziellen Datei verschmolzen sind. So kann eine Datenbank z.B. sowohl Text und audiovisuelles Material beherbergen. Meist wird allerdings zugunsten von Speicherplatzbedarf darauf verzichtet und es werden stattdessen lediglich Verweise auf die externen Daten eingepflegt. Genau dies ist aber für LZA nicht zielführend. Ein digitales Objekt hat als Einheit verstanden und der Nachwelt erhalten zu bleiben. Es darf daher keine Verweise irgendwohin geben, die ins Leere führen.

2.7 Daten, Information, Wissen, Weisheit

In dieser Arbeit geht es um digitale LZA. Digitale LZA heißt allerdings auch die Nutzbarkeit der digitalen Daten zu ermöglichen. Würde man diesen Gesichtspunkt außer Acht lassen, könnte man durch reines Umkopieren des Bitstreams auf jeweils aktuelle Datenträger die Behauptung aufstellen, man hätte hier bereits lz-archiviert. LZA beinhaltet jedoch weitaus mehr strategisches und vorausschauendes Denken hinsichtlich der „Nutzensarchivierung“. Dazu muss man wissen, wie Wissen und Weisheit entsteht, nämlich durch die Interpretation elementarster Strukturen, d.h. also den Bits und Zeichen als Daten. Erst diese Interpretation ermöglicht daraus Informationen abzuleiten und Wissen, sowie Weisheit daraus zu erlangen. Es sei hier auf die einleitend erwähnten Dateiformate hingewiesen, die man ohne Kenntnis, worum es sich handelt, nicht interpretieren könnte. Wissen und Informationsgewinn aus Bitströmen stellt somit nichts anders dar, als Information aus einzelnen Zeichen zu interpretieren, die Worte, dann Sätze, anschließend Sinn und erst letztlich Wissen sowie ableitbare Weisheit ergeben. Es handelt sich also um eine sich immer weiter spezialisierende Pyramide, wie Wissen erlangt und durch Weisheit entsprechende Aktionen gesetzt werden.



Abbildung 3: Die Wissenspyramide (nach Potenzialfinder, 2016)

Die einzelnen Zeichen repräsentieren hier nichts anderes als Daten aus Bits. Durch die Interpretation erhält man dabei von Ebene zu Ebene mehr Information. Kennt man aber nicht einmal die Zeichengröße (8, 16, 32-Bits,...), in denen der Bitstream kodiert ist, bzw. in welchem Format er vorliegt, ist es faktisch nicht möglich, daraus Zeichen und demnach auch entsprechendes Wissen abzuleiten. Und ohne Wissen können folglich auch keinerlei Aktionen gesetzt werden. Bezogen auf digitale Daten hieße dies, keine Ausgabe auf dem Monitor (Text, Film, Grafik), keine Musik oder auch keine ablaufende Software.

2.8 Stand des Wissens

Derzeit existieren für eine Handvoll AV-Daten, die der Mensch wahrnehmen kann, unzählige Dateiformate für entsprechende Digitalisate. Darunter fallen Texte, Klänge, Filme und davon abgewandelt Software samt Computerspielen sowie Datenbanken. Dies spiegelt sich im Sehen (Text, Film, Datenbanken im Sinne von Sichten), Hören (Musik, Spiele) und Fühlen (Spiele mit z.B. Force-Feedback-Joysticks) wider. Was die Sonderform der Software anbelangt, so ist diese als eigentliches Steuerinstrument in Computern oder Consumer-Geräten wie Wiedergabesystemen für CDs, DVDs und Blu Rays oder Smartphones zuständig. Angesichts der vorhin aufgezählten durch die menschlichen Sinne beschränkten Möglichkeiten, nämlich Sehen, Hören und Fühlen, erscheint es geradezu pervers, für Musik unzählige Dateiformate zu haben. Analoges gilt für Grafiken und Videofilme (Sauter, 2009, S. 17.28ff.).

Zu verdanken hat man diese Tatsachen der historischen Entwicklung. Vielfach haben kommerzielle Hersteller die Formate entwickelt und waren nicht bereit sie anderen Hersteller ohne teure Lizenzgebühren zur Verwendung zu überlassen. Heute ist man mit der „Erblast“ dieser Entwicklungen konfrontiert, indem man schon Schwierigkeiten hat, bestimmte Musik, Filme oder Texte auf anderen Endgeräten wiederzugeben, wie z.B. Kindle-Versionen elektronischer Bücher oder MP3-Dateien in älteren Autoradios. Verschärft wird dieser Umstand gerade dann, wenn es an die LZA geht. Wenn schon Schwierigkeiten in der gleichen Zeitepoche bestehen, unterschiedliche digitale Inhalte zwischen Endgeräten auszutauschen und zu nutzen, wie wird sich dies erst in hundert oder tausend Jahren auswirken? Digitale LZA steht daher vor großen Herausforderungen, Digitalisate nicht nur über die Zeit zu retten, sondern auch deren Nutzbarkeit zu ermöglichen. Als weitere Erschwernis kommt die begrenzte Haltbarkeit von Trägermedien hinzu. Letztlich ist man seitens der Archive noch mit immer höheren Datenvolumina konfrontiert, die im Laufe der Zeit produziert werden (Brodersen, Schrimpf, 2009, S.8.12). Bisher waren auch Meinungen vorherrschend, dass man Backups anfertigen sollte, um für den Fall des Datenverlustes darauf vorbereitet zu sein. Wie im Kapitel zum 3-2-1-Backup gezeigt wird, mag dies zwar kurzfristig zutreffen, mit einer LZA-Strategie hat dies allerdings nichts zu tun.

3 Methodik und persönliches Vorgehen

Zur Beantwortung der Forschungsfrage wurde eine systematische Literaturrecherche durchgeführt, wobei der Schwerpunkt auf häufig vorkommenden Autoren liegt. Zentrale Begriffe dabei waren „Langzeitarchivierung“, „Datenträger“, und „OAIS“, die auch für die Recherche in Bibliotheken und Online-Katalogen verwendet wurden.

Im Kapitel zur Digitalisierung als Konvergenz analoger Daten erfolgt die Heranführung an die Problematik der digitalen LZA. Gezeigt werden verschiedene Aspekte, die es in Hinblick auf sog. Digitalisate zu beachten gilt, was auch die Grundlauge für digitale Archivierung darstellt.

Im Anschluss daran erfolgt im Kapitel zu Backups die Erläuterung der Gefahren, die jede Art von Bitstream betreffen kann, was durch mangelnde Konzepte diesbezüglich verursacht wird. Ziel dabei ist es, zu verdeutlichen, wie Backups für Archivierungsstrategien eine zentrale Rolle spielen, obgleich sie selbst nicht zu diesen zählen.

Weiteres erfolgen im Kapitel zu Arten von Daten konkrete Szenarien für die Texte, Filme, sonstiges AV-Material, Software und Datenbanken. In diesem Kapitel liegt der Schwerpunkt darauf, ein Verständnis dafür zu entwickeln, dass es keine standardisierte Vorgehensweise für die LZA gibt. Es werden wichtige LZA-Dateiformate behandelt, wobei es hier immer auf das jeweilige Ziel und den Umfang der geplanten Archivierung ankommt. Das Augenmerk soll besonders auf die sog. Metainformationen gelegt werden. Ohne solche ließen sich archivierte Daten nämlich nur sehr mühsam bis kaum wiederfinden.

Das anschließende Kapitel widmet sich den Datenträgern sowie der Emulation und Migration. Die Möglichkeiten und Grenzen beider Strategien werden hier behandelt.

Im entscheidenden Kapitel über OAIS wird schließlich den internationalen Bestrebungen einheitliche LZA-Strategien zu schaffen, nachgegangen. Dabei stehen nicht individuelle Ausprägungen für spezielle Anwendungsfälle, sondern tatsächlich die anerkannten und üblichen Strategien und weniger die Datenträger im Vordergrund. Nichts ist nämlich schlimmer als wenn unterschiedliche Archive unterschiedliche Strategien zur LZA verfolgen. Das Ziel des Zugriffs und der Lesbarkeit archivierter digitaler Daten in tausenden von Jahren würde hier nämlich nicht erreichbar sein.

Abgegrenzt werden sämtliche Kapitel dadurch, dass sie bewusst nicht mit übermäßigen technischen Spezifikationen an Zahlen, Daten und Fakten gefüllt werden. Es sollen nämlich langfristige Strategien in den Vordergrund gerückt werden und nicht eine Sammlung an Daten von in einigen Jahren ohnehin überholter Technologien erfolgen. Denn zu den Technologien erhält man in zehn bis 20 Sekunden Google-Recherche Material in Hülle und Fülle.

4 Digitalisierung als Konvergenz analoger Daten

Warum ein Kapitel zur Digitalisierung analoger Daten in einer Arbeit, die sich eigentlich mit rein digitaler LZA beschäftigt, unerlässlich ist, wird im Folgenden erklärt. LZA digitaler Daten als eine Strategie verstanden ist nur deswegen erforderlich, da heutige Datenträger eine begrenzte Lebensdauer aufweisen. Einer davon bietet jedoch für geschätzte 500 Jahre zuverlässigen Zugriff auf die Daten und zwar der Mikrofilm. Der Grundgedanke hinter diesem eigentlich analogen Belichtungsdatenträger, der vornehmlich für Dokumente und Bildarchive verwendet wird, ist, dass er auch Daten in Form von Eins und Null als in Binärcode gebrachte Druckversion, ähnlich einer Seite Papier, aufnehmen kann. Damit kann ein Bitstream ohne Qualitätsverlust erhalten werden. Der Mikrofilm kann somit gleichsam analoge Dokumente als Fotografie speichern wie auch moderne Blu Ray-Filme als Bitfolge. Wie einleitend allerdings das Problem aufgeworfen wurde, lässt sich damit ohne Kenntnis des Daten- und Dateiformats sowie der entsprechenden Verarbeitungsanlagen, welche diese interpretieren, nicht viel anfangen. Dies ist aber nicht das Problem des Trägermediums Mikrofilm, sondern ein Problem des Redigitalisierens – dies müsste für einen Blu Ray-Bitstream eben ein Hard- oder Software-Blu Ray-Player übernehmen. Für den Mikrofilm gilt jedoch, dass zu seiner Lesbarkeit lediglich Licht und entsprechende Vergrößerungslinsen, in Zukunft entsprechend dafür optimierte Scanner, die wiederum einen digitalen Bitstream für Computer daraus aufbereiten, erforderlich sind. Da es sich also um einen nicht ausschließlich kompliziert technisch lesbaren Datenträger handelt, erfreut er sich als eine Art Hybrid großer Beliebtheit. Auf den Punkt gebracht wird eine Analog/Digital und Digital/Analog-Rückumwandlung vollzogen. (Keitel, 2009, S.8.33). Wie im Kapitel zu Arten von Daten noch gezeigt wird, kommt es aber letztlich immer darauf an, was konkret archiviert werden soll.

Analoge Medien wie Bücher aus säurehaltigem Papier, die sonst zerfallen würden (Bárányi, 2006, S.66f.), lassen sich im Zuge der LZA digitalisieren, um dadurch Vorteile zu nutzen. Zu nennen wären einfachere Vervielfältigung einer Digitalkopie, schnelle Durchsuchbarkeit und einfacherer Zugriff. Leider gibt es auch Nachteile, welche durch sog. Digitalisate entstehen. Digitalisierung ist nämlich immer mit Informationsverlust verbunden. Besonders audiovisuelles Material leidet unter systembedingter Datenreduktion, da die Scanoptik für Fotos begrenzt, bzw. die Samplingrate für Musik Zwischenstufen verloren gehen lässt. Bei Büchern kann zwar der reine Text erfasst, aber niemals die Authentizität eines gedruckten Werkes durch eine Digitalisierung entstehen. Bezogen auf das OAIIS-Modell, welches als anerkanntes Referenzmodell zur Archivierung dient, spielt Digitalisierung analoger Quellen jedoch keine Rolle. OAIIS geht hier grundsätzlich ausschließlich von bereits digitalen Daten aus. Wie von Suchodoletz treffend dazu feststellt, hängt der Erfolg der LZA analoger Medien allerdings vom langfristigen Zugang zum digitalen Archiv ab.

Anzumerken wäre, dass ein Großteil der Bestände in Bibliotheken und Archiven immer noch ausschließlich rein analog vorliegt (von Suchodoletz, 2009, S.45f.).

Genau vor diesem Hintergrund, dass durch Technologiewandel bedingt, kaum mehr analoge Speicher bei der Produktion von Daten verwendet und auch kaum mehr produziert werden, wäre die Suche nach einer Strategie für digitale LZA unvollständig, wenn man die Probleme außer Acht ließe, die eine Digitalisierung analoger Informationen aus der Wirklichkeit mit sich bringt. Grundsätzlich gilt dabei, die Entscheidung darüber zu treffen, welches Ausgangsmaterial vorliegt. Die vorhin erwähnten Bücher bringen z.B. das Problem mit sich, dass man schon auf Schwierigkeiten stößt, sobald auch nur eine einzige Illustration darin enthalten ist. Das beste Beispiel hierzu wären Magazine oder einfache Zeitungen, die auch Bilder enthalten. Nun könnte man einwenden, dass man Bilder einfach grafisch scannen könnte. Man vergisst hier allerdings, dass in diesem Fall die Durchsuchbarkeit darunter leiden würde, sofern man nicht den Text noch zusätzlich per OCR separat scannt. Auf diese Weise erhält man bereits zwei Typen von Digitalisaten, nämlich den abstrahierten Text und die separierten Grafiken. Gefragt sind also Mechanismen, die der Durchsuchung und Zusammenführung von Digitalisaten dienen – egal ob Text, Bilder, Ton, Programme, Videos, etc... Hughes bringt die Problematik treffend auf den Punkt:

“As with all media types, appropriate metadata will be essential to guaranteeing the creation of a ‚well-formed digital object that will survive for the long term‘, as well as access to the material. All data must be documented properly to ensure that future curators and users understand the materials that they are dealing with. The creator of digital objects must expend almost as much effort on the development of appropriate metadata as on the digitization itself.” (Hughes, 2004, S.248)

Hier wird also der Begriff des digitalen Objektes umschrieben. Genau ein solches digitales Objekt macht dabei auch Sinn. Kapselung von zusammengehörigen Daten als Information in einem Objekt erleichtert nämlich die Auffindbarkeit zusammengehöriger Informationen. In Kombination mit obiger Feststellung Suchodoletz‘ hinsichtlich des Vorliegens großer Mengen an Beständen analogem Quellmaterials begegnet man in der LZA gleich mehreren Herausforderungen. So müssen rein digital produzierte Inhalte zwecks Auffindbarkeit kompatibel zu denjenigen sein, die erst nachträglich aus analogen Daten gewonnen werden. Unterschiedliche Quelldaten müssen in einer Art Medienkonvergenz zusammengeführt werden und Aufgabe eines langzeitstabilen Dateiformates ist es neben diesen Merkmalen die Auffindbarkeit zu gewährleisten. Genau dies leisten Metadaten, d.h. also vereinfacht ausgedrückt Daten über Daten anzulegen. Im Optimalfall geschieht dies gleich bei deren Erstellung, spätestens aber sollte es bei der Aufnahme in

ein LZ-Archiv erfolgen, dem sog. Ingest, wie im Kapitel zu OAIS noch gezeigt wird.

Für die spätere digitale LZA ist besonders von Bedeutung, bereits bei der Erzeugung von Digitalisaten auf korrekte Dateiformate zu achten. Um hier zu entscheiden, welche diese sind, muss man darüber nachdenken, welches Material man eigentlich lz-archivieren möchte. Lazinger bringt dies auf den Punkt, indem sie darauf verweist, dass Entscheidungen über die Archivierung, egal ob noch analoges oder schon digitales Ausgangsmaterial vorliegt, immer auch eine Entscheidung darüber sind, in welchem Umfang dies geschieht. Der Hintergrund hierbei sind Querverweise. Als Beispiel mag ein Buch dienen, welches ein Literaturverzeichnis enthält. Das Buch wäre nun nicht mehr vollwertig, wenn die darin angegebenen Quellen nicht ebenfalls irgendwo archiviert werden würden. Auch archivierte Internet-Seiten mit ins Leere verweisenden Hyperlinks wären genauso wenig wert, wie einzeln archivierte Episoden von TV-Serien oder mehrteiligen Filmen. Zudem stellt sich auch hier die Frage nach der Menge an Metadaten. Das Problem dabei ist, dass multimediale Dokumente auch gemischten Content enthalten können und sich daher nicht einfach zu einem Objekt kapseln lassen, wie es weiter oben Hughes im Sinn hatte. Werden hier nämlich zu jedem Teilbereich Metainformationen angelegt, wäre eine zielführende Suche ebenfalls nicht möglich, da man dann erst recht den gesamten Content manuell durchsuchen müsste, auch wenn man nur ein einzelnes großes digital archivierte Gesamtobjekt erhält. Die Dynamik sich ständig ändernder und wachsender Inhalte lässt sich dabei nur durch sog. Gesamt-Snapshots archivieren (Lazinger, 2001, S.17f.). Dabei taucht aber ein neues Problem neben den Dateiformaten auf: der Speicherplatzbedarf.

Hier kann man zwar Entscheidungen treffen, verlustbehaftet zugunsten weniger Speicherplatzbedarfes oder ohne Qualitätsverlust, dann aber zu Lasten von höherem Bedarf zu digitalisieren. So kann z.B. Musik mit 8 Bit und 22 khz abgetastet werden oder mit dem vollen menschlichen Klangspektrum von 16 Bit und 44, bzw. 48 khz (doppelte Abtastrate von 22-24 khz). Während das eine Telefonqualität hat, hat das andere CD-Qualität. Eine spätere digitale LZA mag an erzeugter schlechter Ausgangsqualität nämlich nichts mehr zu ändern. Eine Entscheidung für das Dateiformat bestimmt hier bei der Erzeugung auch wesentlich das weitere Schicksal des Digitalisates im Sinne der LZA. Ähnliches gilt für Fotos und Filme, die man mit zu geringen Auflösungen oder Bildwiederholfrequenzen abtastet, bzw. manuell nachbessern muss, was immer mit individuellen Fähigkeiten von Personen verbunden ist (Black-Veldtrup, Meusch, Przigoda, 2002, S.112). An dieser Stelle wird einmal mehr deutlich, dass man analogen Eingangsdaten doch mehr Beachtung hinsichtlich der Qualität bezogen auf spätere digitale LZA schenken sollte. Was gebe man heute dafür, hätte man schon zu Zeiten von Elvis Presley hochwertige A/D-Wandler und 7.1-Surround-Sound sowie HD-Auflösung gehabt?

5 Backup ist keine Archivierung – Unterschiede

„Aber ich hab doch eh ein Backup gemacht“ ... Diesen Satz hört man sehr oft, wenn jemand Daten trotz vorhandenem Backup verloren hat. Dies passiert z.B. schon dann, wenn auf fehlerhafte Datenträger gesichert wird, ohne sich durch Verifikation davon zu überzeugen, dass die Daten auch wieder eingelesen werden können. Ein Backup ist eben *kein* Backup. Bezogen auf LZA gilt dieser Satz umso mehr, denn ein einziges Archiv kann keine zuverlässige LZA-Strategie gewährleisten, da immer Unfälle in Form von Naturkatastrophen, technischem oder menschlichem Versagen drohen und dann der einzige archivierte Datenbestand vernichtet wäre. Ein Verständnis von Backup-Strategien kann hier die LZA perfekt unterstützen und ergänzen. Backups dienen v.a. dazu, Zugriff auf Daten im Optimalfall im 24/7-Betrieb, also rund um die Uhr zu ermöglichen und Systemausfällen vorzubeugen. Im Falle des Falles besteht das Ziel darin, Systemausfallszeiten zu minimieren. Damit ist zum Ausdruck gebracht, aktuelle und kritische Daten auf schnellen Speichern verfügbar zu halten. Hier wäre z.B. an Kundendatenbanken oder zur Erfüllung laufender Aufträge vorhandener Multimedia-Daten zu denken, wie z.B. in der Webentwicklung.

Archivierte Daten werden dagegen idR. nicht permanent gebraucht und müssen daher auch nicht auf schnellen Speichern liegen. Primärer Zweck jeglichen Archives ist der Datenerhalt und nicht die schnellstmögliche Verfügbarkeit. Im Optimalfall wäre das zwar so, in der Praxis macht die Knappheit von Ressourcen wie eben dem bereits erwähnten Speicherplatzbedarf sowie der Betrieb von Hochverfügbarkeitsanlagen und damit letztlich das Budget diesem Unterfangen einen Strich durch die Rechnung. Gefragt sind daher Strategien, mit möglichst geringem Ressourcenbedarf ein Optimum für LZA-Strategien herauszuholen. Gleichzeitig soll redundantes Vorhandensein von Datenbeständen gewährleistet sein, für den Fall, dass einer davon doch untergeht – aus welchen Gründen auch immer.

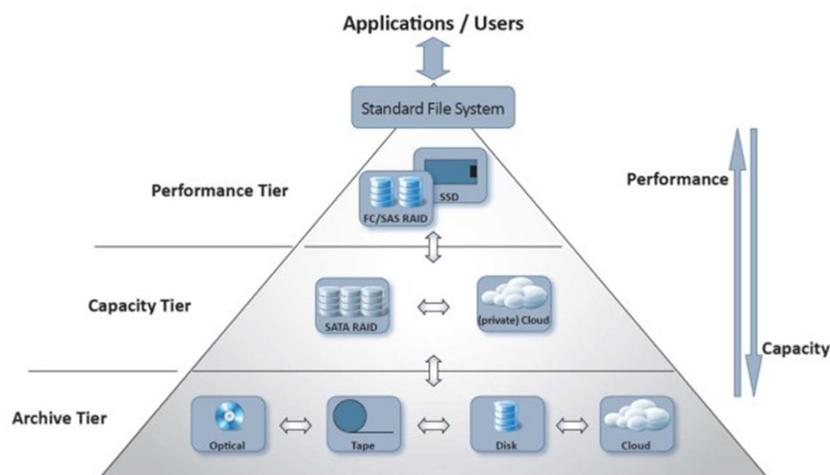


Abbildung 4: Schnelle und langsame / geringe und hoch kapazitative Speicher (Rieder, 2009)

Die Abbildung zeigt hier deutlich, wofür man schnellere Speichersysteme benötigt und wofür langsamere, kostengünstigere völlig ausreichend sind. Jedes Archiv muss hier allerdings individuell entscheiden, was es ständig abrufbar bereit hält und was erst „ausgehoben“ werden müsste.

5.1 Das 3-2-1-Backup-Schema

Um möglichst ressourcenschonend Daten zu sichern, ist jedenfalls die sog. 3-2-1-Regel von Bedeutung. Bei dieser Strategie werden mind. drei Datenbestände vorrätig gehalten, nämlich primär die Original-Datenbestände auf den Verarbeitungssystemen selbst sowie mindestens zwei Backups davon. Diese Backups müssen dabei durch zwei verschiedene Technologien angefertigt sein. Empfohlen hat sich dabei zudem die Vorgehensweise, einen Backup-Satz in einem feuerfesten Bereich oder zumindest in einem anderen Brandabschnitt zu lagern, den anderen fernab außer Haus. Daher der Name 3-2-1 (3x Daten, 2x auf verschiedenen Technologien, 1x außer Haus). Liegen die Originaldaten selbst noch RAID-gesichert auf entsprechenden Verarbeitungssystemen, d.h. entweder gespiegelt oder mittels Paritätsprüfung, so muss schon sehr viel schief laufen, um hier Daten zu verlieren. Man darf allerdings nicht den Fehler machen und RAID mit einem Backup verwechseln, da es nur gegen plötzlichen Datenverlust, nicht aber gegen Viren, bewusstes Löschen oder vor fehlerhaften Dateneingaben schützt. Derartige Aktionen würden sich nämlich auf den gesamten RAID-Verbund auswirken. Im folgenden Bild repräsentiert die Feuerfestigkeit der Safe, das Lagern außer Haus das Bankschließfach sowie die unterschiedlichen Backup-Strategien das NAS-System und die Backup-Bänder. Freilich kommen in LZ-Archiven entsprechende Datenträger mit höherer Lebensdauer zum Einsatz. Erinnerung sei hier daran, dass es sich um eine Backup-Strategie und nicht um eine LZA-Strategie handelt. Sie bildet allerdings die Grundlage für LZ-Archive.

Generell gilt hier z.B. seitens des deutschen Bundesarchives die Empfehlung nur weit verbreitete Datenträger zu verwenden für die internationale Standards gelten. Praktiziert werden dort die Strategien Umkopieren auf die jeweils aktuellen Datenträger der jeweiligen Zeit, wobei davon immer zwei verschiedene Technologien verwendet werden, also z.B. Band sowie optische Datenträger. Wichtig ist auch deren getrennte Lagerung. Somit wird auch hier deutlich, dass das erwähnte 3-2-1-Backup eine bewährte Strategie zur Archivierung ist. Bis heute trat nämlich beim deutschen Bundesarchiv kein Datenverlust auf (Rathje, 2002, S.119f.) Wie die folgende Abbildung zeigt, stellt die Strategie schematisch dar, dass man sich mittels NAS-System als schnellem Speicher sowie Bändern als langsamen Archivmedien mit dem 3-2-1-Backup auch in gewissem Umfang ein eigenes Archiv aufbauen kann. Den Server muss man sich dabei natürlich RAID-gesichert vorstellen. Damit hat man die Daten dreimal, nämlich RAID, NAS und Band, zweimal auf verschiedenen Speichern (Band, NAS) und einmal outgesourced außer Haus im Bankschließfach (=3-2-1)

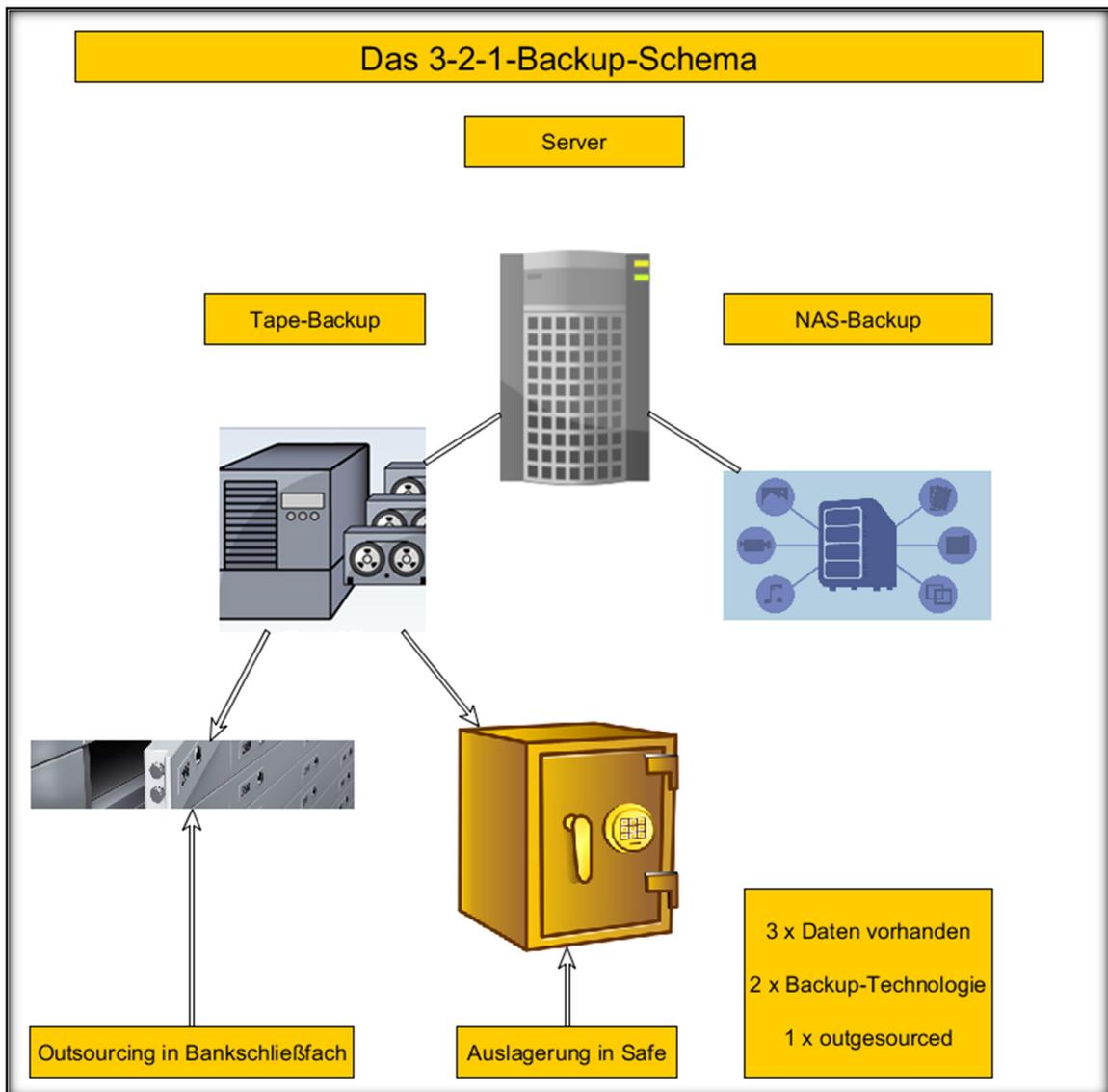


Abbildung 5: Das 3-2-1-Backup (Eigenzeichnung mit yed / Einzelbildquellen im Anhang)

5.2 Das Generationen-Backup

Eine Backup-Strategie ist allerdings nur so gut, wie häufig man sie auch anwendet. Führt man nur einmal ein Vollbackup durch und danach nie wieder, so wird man sicherlich auch irgendwann einmal Daten verlieren. Selbiges gilt, wenn man zwar zusätzlich täglich inkrementelle Backups durchführt, aber mangels Organisation einige Backup-Sätze davon verloren gehen sollten. Vielfach wird daher am Wochenende ein Vollbackup gefahren, während unter der Woche differentielle Backups gemacht werden. Ergänzt werden diese Strategien dann durch allmonatliche Monatsbackups sowie jährlich mit einem Jahresbackup. Es handelt sich hier um das sog. Generationenprinzip Vater / Großvaterbackup (Luther, 2004). Würde man für tägliche Backups auf zwei Systemen 730 Mediensätze benötigen, benötigt man mit der richtigen Strategie für einen Restore, der bis zu einem Jahr zurückliegt, nur 36 Mediensätze. Folgende Strategie hat sich dabei in vielen Fällen für KMUs bewährt:

Backup wann	Anzahl Backup-Mediensätze	Art / wie oft
Mo-Do	4	differentiell / 52x im Jahr überschrieben, sollten daher häufig ausgetauscht werden
Fr	1	Vollbackup / bis zu 13x im Jahr überschrieben
Monatsende	12	Vollbackup / nur 1x im Jahr überschrieben
Jahresende	1	Vollbackup / nur 1x beschrieben
Summe	18 (= 36 für 2 Technologien)	

Tabelle 2: Restore mit wenigen Mediensätze (eigene Tabelle)

Hinsichtlich der Backup-Strategien selbst führt das differentielle Backup Sicherungen nur geänderter Daten seit der letzten Sicherung durch, was daher auch schneller erledigt ist. Folglich laufen in der Praxis differentielle Backup-Jobs meist in einer Nacht durch, während am Wochenende die Vollbackups in Ruhe bis Montag durchlaufen können. Inkrementelle Backups unter der Woche sind weniger zu empfehlen, da man alle Zwischenstufen der Mediensätze zur Wiederherstellung benötigen würde. Mit obiger Strategie dagegen wird mit vergleichsweise günstigen Mitteln ein Maximum an Effizienz beim Backup erzielt. Lediglich bei Ausfall der Hardware muss dafür Sorge getragen werden, die entsprechenden Komponenten schnellstmöglich wieder zu ersetzen, da man inzwischen nur 2 Datensätze (1x am RAID, 1x auf dem Mediensatz der einen, aber nicht der anderen Technologie hätte). Selbiges gilt für RAID-Verbünde bei Ausfall von Festplatten und / oder Spare-Laufwerken. Auch diese müssen schnellstmöglich ersetzt werden. Sich über ausgefallene Komponenten informieren lassen kann man sich dabei durch Monitoring-Systeme wie z.B. Zabbix oder Nagios / Icinga (Betz, Widhalm, 2016, S.87ff.).

6 Arten von Daten

Digitalen Daten liegen binär als Null oder Eins vor, welche die Zustände Strom fließt oder eben nicht in Datenverarbeitungsanlagen repräsentieren (=Bitstream). Auch Unterhaltungselektronik und letztlich auch Archivierungssysteme basieren auf diesen beiden Zuständen, wobei an CDs, DVDs, bzw. Blu Rays als Datenträger oder verschiedene Neuentwicklungen zu denken wäre. Ein Datenträger kann ganz seinem Namen gerecht werdend nur die Daten „tragen“. Wie lange das der Fall ist hängt von seiner Pflege, dem Verschleiß durch mehr oder weniger häufigen Gebrauch und von den Umweltbedingungen ab. Man merkt aber, dass es hier nicht auf die Zustände Null und Eins ankommt, sondern auf deren korrekter Interpretation. Insofern ist es durchaus legitim, von Arten von Daten oder zielführender ausgedrückt, der durch sie repräsentierten Dateiformate zu sprechen.

Wie einleitend bereits darauf hingewiesen wurde, können digitale Daten als Befehle, Werte oder Zeichen interpretiert werden. Um diess Problematik noch zu vergrößern, muss man wissen, dass Daten in Bytes organisiert sind, welche aus 8 Bits bestehen. Damit lassen sich 256 Zustände darstellen (2^8 , also der Wertebereich von 0-255). Um Werte größer 255 darzustellen, muss man diese Bytes schon zu Words, Double-Words, Q-Words, etc... zusammenfassen. Dann lassen sich Wertigkeiten bis 2^2 hoch der jeweiligen Bitanzahl darstellen. Leider liegt genau darin das Problem, da man dem reinen Bitstrom nicht ansieht, ob er nun ein Zeichen als Byte oder eine Zahl enthält. Werte könnten zudem grafische Pixel darstellen, aber auch Sinuskurvenwerte für Musik. Auch Videoinformation wäre denkbar. Bezogen auf die LZA heißt dies, dass man zwar die langlebigsten Datenträger verwenden kann, um für tausende von Jahren für die Datenarchivierung gerüstet zu sein, jedoch kann ohne ein sekundäres Archiv über die Datenstrukturen, d.h. Metainformationen, keine Aussage über den Inhalt getroffen werden.

6.1 Interpretationsregeln

Um die Problematiken der unterschiedlichen Darstellungsweisen und Kodierungen zu umgehen, wurden eindeutige Interpretationsregeln geschaffen. Diese liegen im Dateisystem sowie -format verankert. Während Dateisysteme hier die logische Organisation der einzelnen Dateien definieren, definieren die Dateiformate die in ihr gespeicherte Information selbst. Der Grund für diese Trennung liegt wiederum im Bitstream begründet. Ohne Kenntnis des Startbereiches einer Datei, kann man auch hier keinerlei Aussagen über die Informationen machen. Man stelle sich hierzu vor eine Textdatei eines Datenträgers würde fehlerhaft gelesen werden und zwar nicht von Beginn an, sondern mitten drin. Der Sinngehalt wäre ein völlig anderer. Somit stellt sich hier die Frage nach langzeitstabilen Dateisystemen einerseits und -formaten andererseits.

Was Dateisysteme anbelangt kann man hier das weit verbreitete NTFS des Herstellers Microsoft nicht als lza-tauglich einstufen. Erstens sind viele Strukturen nach wie vor undokumentiert, was nicht zuletzt den Sicherheitsbedenken des Herstellers Microsoft entspricht, hier so wenig Dokumentation wie möglich darüber preiszugeben, andererseits sind Open Source Betriebssysteme, welche es lesen können nicht zu 100% dazu kompatibel. Einer LZA steht hier der Gedanke der freien Verfügbarkeit und quelloffener Dokumentation entgegen. Somit kommt als Betriebssystem für die Dateisysteme nur Linux infrage. Dies ist v.a. dann der Fall, wenn es um Archivierung von Big Data geht, was sich mit Hadoop auf verteilten Rechenzentren, bzw. verteilten Rechnern auf der ganzen Welt, den sog. Nodes bewerkstelligen ließe. Exemplarisch mag dafür Facebook dienen:

“As a result, the cluster size and usage grew by leaps and bounds, and today Facebook is running the second-largest Hadoop cluster in the world. As of this writing, we hold more than 2PB of data in Hadoop and load more than 10 TB of data into it every day. Our Hadoop instance has 2,400 cores and about 9 TB of memory, and runs at 100% utilization at many points during the day. We are able to scale out this cluster rapidly in response to our growth, and we have been able to take advantage of open source by modifying Hadoop where required to suit our needs.” (White, 2012, S.557)

Hierzu existieren zwar Archivierungsmöglichkeiten wie sie auch von Google und Youtube mit den Rechenzentren vorgemacht werden, allerdings handelt es sich hierbei um elektronische Speicherung. Man stelle sich vor, die beinahe täglichen Änderungen der Datenbestände müssten gesichert, katalogisiert und dann in Archiven physisch aufbewahrt werden. Alleine der logistische Aufwand hierzu wäre kaum zu bewältigen. Rechenzentren sind in Zeiten von Hackern und globaler Vernetzung allerdings natürlich sehr fragile Orte für digitale LZA. Damit stellt sich die Frage, ob denn derartige Daten in hunderten von Jahren noch zugänglich sein werden. Derzeit mangelt es dafür allerdings noch an Offline-Datenträgern mit ausreichender Kapazität für die LZA.

6.2 Langzeitstabile Dateiformate

Verfügt man über digitale Daten, fehlen einem nach wie vor lza-taugliche Dateiformate. Hier gilt grundsätzlich das Gleiche wie bei den Dateisystemen. Das Format hat quelloffen und nicht proprietär zu sein, zudem eine hohe Verbreitung aufzuweisen und muss frei verfügbare Betrachtungsprogramme bieten (Landesarchiv Nordrhein-Westfalen, 2015, S.8). Ein wichtiges Format in diesem Zusammenhang ist PDF/a. Dieses Format ermöglicht die LZA von Text, Vektorgrafik und gerasterten Bildern (PDF Tools AG, 2009).

6.3 PDF/a (Texte, Rastergrafiken, Bilder)

Da man in Unkenntnis eines Dateiformates einer Archivdatei keinen Informationsgewinn aus einem Bitstream ablesen kann, nahm man sich diesem

Problem im Jahre 2005 bei der International Standards Organization (ISO) an. Als Ergebnis wurde das PDF/a-Format präsentiert. Dessen Vorteile sind:

1. strukturierte Objekte, wie Text, Vektor- oder Rastergraphiken mit Unterstützung des TIFF-Formates, welches hier per Texterkennung im Vorfeld allerdings erst aufbereitet werden muss.
2. Kompression, da hier nur ein Bruchteil des Speicherplatzes einer TIFF-Datei benötigt wird.
3. Metadaten wie Autor, Thema, Inhalt, Schlüsselwörter, Erzeugungs- und Änderungsdatum, Herausgeber, etc... werden LZA-Konform unterstützt.
4. keine spezielle Bindung an ein spezifisches Ausgabegerät. Konkret heißt dies, dass erst zum Zeitpunkt der Darstellung das Dokument entsprechend gerendert wird.

Bedenklich erscheint in Hinblick auf LZA bei PDF/a allerdings, dass Hersteller Adobe innerhalb von 15 Jahren bereits acht Revisionen herausgegeben hat. Damit wollte man neue Entwicklungen unterstützen. Bezogen auf die LZA heißt dies aber einmal mehr, dass man mit unterschiedlichen Dateiformaten konfrontiert ist. Als gemeinsame Basis von PDF/a ist dabei Referenz 1.4 anzusehen, wo Transparenz, Integration multimedialer Inhalt wie Audio und Video jedoch nicht erlaubt sind. Stattdessen müssen alle verwendeten Schriftarten eingebettet sein, was letztlich auch Sinn macht, da das Dokument ansonsten nicht korrekt in hunderten von Jahren dargestellt werden könnte, wenn das digitale Objekt nicht vollständig gekapselt wäre. Das oben genannte TIFF-Format als frühe Form von PDF bis zur Übernahme des Entwicklers Aldus durch Adobe ist allerdings nach wie vor ebenfalls Standard in der LZA und wird in der Praxis weiterhin eingesetzt (Enders, 2009, S.17.11)

Mit PDF/a ist jedoch nicht das Allheilmittel gefunden. Zudem existiert davon eine Basic-Version, die exakte reproduzierbare Darstellbarkeit gewährleistet, während die vollständige A-Version zudem die komplette Abbildung des UNICODES mit eingebetteten Schriften und barrierefreien Zugang über Screenreader ermöglicht. Wie Ludwig treffend feststellt, können sich Ansprüche auch an einmal als die beste Wahl erscheinende Dateiformate ändern oder technischer Fortschritt bringt neue Dateiformate hervor, welche auch dort Änderungen erforderlich machen. Eine dauerhafte Festlegung auf ein Format für die LZA ist daher nicht möglich (Ludwig, 2009, S.7.9). Zudem verfügt man mit PDF/a zwar über ein Dateiformat - Software, Audio- und Videomaterial lässt sich damit jedoch nicht archivieren, sieht man von einem archivierten „Ausdruck“ des Bitstreams als Binärcode ab. Grundsätzlich gilt aber auch hier wie bei der Mikroverfilmung, dass es machbar wäre. Mühsamer wäre allerdings die Rückführung in das Ausgangsformat als sog. Redigitalisierung (Keitel, 2009, S.8.33). Zudem belegen Audio- und Videodaten schon ein Vielfaches an Speicherplatz, den reine Textdokumente beanspruchen, wobei von HD-Aufnahmen dabei noch gar nicht die Rede ist. Kommen nun noch die ausgedruckten Binärcores als PDF/a dazu, steigt der Speicherplatz nochmals dafür an. Man könnte wenigstens den Binärcode sedezimal (besser bekannt unter dem Namen Hexadezimalsystem) „ausdrucken“, wo eine Ziffer 4 Bits repräsentiert und so zumindest Kompressionsraten um den Faktor vier

erreichen. Bei jeder sonstigen Konvertierung, wie eben auch in ein PDF/a-Dokument ist aber immer darauf zu achten, ob der Vorgang ohne Fehler am Layout und Text erfolgt ist, d.h. also wie beim 3-2-1-Backup also mittels Verifikation. Die Qualität des Ergebnisses hängt dabei vom ursprünglichen Dateiformat und der Software ab, die die Konvertierung durchführt.

6.4 Software

Für Software kann man das Kapitel recht kurz halten. Software ist immer auf ihre jeweilige physische Umgebung abgestimmt und läuft daher auch nur dort - von Emulation einmal abgesehen. Daher gibt es auch keine langzeitstabilen Formate für Software. Es gibt allenfalls virtualisierte Container für die verschiedensten Emulationsumgebungen. An ein Original reicht dies allerdings weder bei Hyper-V, Virtual Box oder VMware heran, den derzeit weit verbreiteten Herstellern vor Virtualisierungs-Software.

6.5 Videoformate

Mit PDF/a hat man ein Format für statische Informationen. Für Laufbilder steht man dagegen weiterhin vor dem Problem der LZA. Wie bereits erläutert wurde, ist es natürlich möglich, den Bitstream dieser Dateien als binäre Folge von Nullen und Einsen in einem PDF/a-Dokument zu archivieren. So ließe sich auch Software „ausdrucken“. Das wäre zwar möglich, aber nicht sehr praktikabel. Beim Landesarchiv Nordrhein Westfalen weiß man um diesen Umstand Bescheid und gibt im „Leitfaden für langzeitstabile Datenformate in der elektronischen Aktenführung“ von 2015 auch zu, dass de facto bis heute noch keine sicheren LZA-Strategien für Audio- und Videoarchivierung existieren. An Problemen werden für ein DMS diesbezüglich der Speicherbedarf und damit einhergehend die nicht permanente Verwaltbarkeit genannt. Empfohlen wird daher, derartige Daten grundsätzlich zu vermeiden oder zumindest über Verweise im Sinne einer hybriden Aktenführung zu setzen. Damit sind freilich auch hier keine Lösungen für die LZA von Audio- und Video genannt (Landesarchiv Nordrhein-Westfalen, 2015, S.13).

Im Klartext heißt dies, dass man AV-Daten heute noch nicht lz-archivieren kann. Natürlich hatte man bei der Behörde eher verwaltungsrechtliche Verfahren im Sinn, doch ändert dies nichts an der Problematik der fehlenden Strategien dafür. Es existieren allenfalls Behelfslösungen wie das rechtzeitige Umkopieren auf neuere Datenträger. Big Player im Web wie Google mit Youtube oder Facebook machen es allerdings vor, dass man mit genügend Geldmitteln verteilte Rechenzentren betreiben kann, um hier elektronische Archivierung durchzuführen, wie es bereits mittels Hadoop erwähnt wurde. Der Auffassung des Landesarchives Nordrhein Westfalen steht allerdings die DIN-Norm 31644 in Kriterium 23 entgegen, wo für digitale Langzeitarchivierung von Videomaterial das MPEG4-Format als tragendes Format für Inhaltsdaten neben den eigentlichen Container-Paketen genannt wird (Keitel, Schoger, 2013, S.105). Anzumerken ist dazu, dass es sich dabei

auch um jenes Format der DVD-Filme handelt. Verwendet man hier die M-Disc, die weiter unten angesprochen wird, kann man allerdings von LZA ausgehen.

6.6 Metainformationen

Während Primärinformationen den Bitstream der jeweiligen Datei selbst repräsentieren, handelt es sich bei Metainformationen um zusätzliche, diese beschreibende Informationen. Das PDF/a-Format z.B. unterstützt als langzeitstabiles Dateiformat bereits Metadaten. Um diese Informationen verwerten zu können, liegen diese in einem DMS fast immer in Form einer Datenbank vor. Hierbei ist es nicht problematisch, wie diese Datenbank gestaltet ist, denn sollte die Datenbank verloren gehen, lässt sich diese jederzeit anhand des neuerlichen Einlesens der betreffenden Dokumente neu erzeugen - es kostet nur Zeit. Problematischer ist allerdings, dass im Falle des Austausches der Dokumente keine automatische Konvertierung in das Austauschformat XDOMEA2 erfolgt. Somit ist man faktisch dazu gezwungen, den Einlesevorgang für jedes individuelle DMS zu wiederholen.

Wesentlich dabei ist, die Standardschemata der Dokumente nicht durch individuelle Ergänzungen zu modifizieren und falls doch, diese Informationen nur als Plain Text den standardisierten Metainformationen hinzuzufügen. Anderenfalls kann das passieren, was in Datenbanken zu den größten Unsitten gehört, nämlich keine atomaren Daten vorliegen zu haben, sondern als „Autor“ z.B. die betreffende Person einmal in der Form „Vorname Nachname“ anzugeben, wogegen andere Personen diese Metainformationen als „Nachname Vorname“ einpflegen. Für eine Datenbank eines DMS ist ein(e) Josef(a) Müller eben nicht dieselbe Person wie ein(e) Müller Josf(a). Hier ist also an die Disziplin der Archivare zu appellieren, aber auch an einheitliche Standards, die es bis dato nicht gibt. PDF/a als Dateiformat sieht hier nämlich auch keinen atomaren Vor und Nachnamen sondern nur „den Autor“ vor.

Dass v.a. Dublin Core Metadaten eine große Rolle spielen, davon zeugt, dass niemand geringerer als die digitale Strategie der US Library of Congress auf diesen Informationen basiert (Library of Congress, 2000, S.135). Dieser Standard, genannt METS findet auch in modernen XML-Dokumenten Anwendung, die in Datenbanken abgebildet werden können, was ein richtiger Schritt zur LZA ist. Erinnerung sei hier an die im Kapitel zu Arten von Daten erwähnten proprietäre Datenbanksysteme, die kein einheitliches Format aufweisen. Verwaltet wird METS dabei von der US Library of Congress (Egger, 2005, S.42). Um diese Daten herum lassen sich dann von Archiv zu Archiv eigene, spezielle Use Case Szenarien implementieren. Konkret greift man hier auf einen einheitlichen Datenstandard zu, schafft sich allerdings das eigene Framework für die Anpassung an lokale Gegebenheiten und ohne Gefahr zu laufen, Metadaten zu verfälschen (Egger, 2005, S47ff).

Im Zusammenhang mit diesen Problemen ist auch von strukturierter Konvertierung die Rede, welche entsprechenden Aufbereitungsregeln hinsichtlich Layout, Zeichen, Struktur und Bildern folgt, um hier gleich automatisch Metadaten zu erfassen, bzw. doppelseitig gescannte Textdokumente automatisch zu separieren. Bei 100.000 und mehr Dokumenten, die man digitalisiert, wäre nämlich die manuelle Metadatenerfassung nicht zielführend. (Gravenhorst, 2007, S.178ff.). Auch die routinemäßige Überprüfung derart automatisch erfasster und mehreren Standards unterworfenen Metadaten ist Teil der LZA-Strategie. An Standards kommen dabei JSTOR, XML und METS bei der deutschen Nationalbibliothek zum Einsatz. (Brodersen, Schrempf, 2009, S.18.9.f) Anzumerken ist dazu, dass in der Praxis zerknitterten Papierstellen, Personalnot, Staub- und Kaffeeflecken, etc... die automatische Generierung von Metadaten behindern können sowie der Unwille kleinerer Archive, die hohen Standards an Metainformationen zu erfüllen.

Problematisch wird dies, wenn, wie die US-Library of Congress z.B. für ihre Partner als sog. Fail-Safe-Mechanismus dient. Der dahinter steckende Gedanke ist nämlich, dass zwar grundsätzlich jedes Archiv (=Bibliothek im Sinne einer Gedächtnisorganisation) ihr Archivgut selbstständig verwaltet. In dem Moment aber, wo dies nicht mehr möglich ist aufgrund von Kapazitätsengpässen, Personalnot, bevorstehenden absehbaren Naturkatastrophen, etc... eben die Library of Congress einspringt. Dabei steht nicht notwendigerweise im Vordergrund Server der Library of Congress zu verwenden, sondern stattdessen verteilte Systeme ähnlich der Rechenzentren von Facebook und Google, was letztlich nichts anderes als Big Data bedeutet. Ohne einheitliche Formate ist man an dieser Stelle allerdings wieder mit Konvertierung, Migration und allenfalls Emulation beschäftigt. Evaluierungskommissionen bemängelten hier bereits das Fehlen der technischen Standards dafür (Library of Congress, 2000, S.110f.)

Neben Dublin Core existieren noch andere Standards im Bereich der Metadaten. Hervorzuheben wäre noch PREMIS, das auf fünf Teilaspekten basiert (Brandt, 2009, S.6.11f.):

- **Intellectual Entities:** Diese beschreiben das Objekt nach Art eines Inhaltsverzeichniseintrages, können allerdings auch Querverweise enthalten.
- **Object Entity:** Diese beschreibt das Objekt selbst, die Struktur und den Aufbau für ein LZ-Archiv.
- **Events Entity:** Dies ist vergleichbar einem Logfile. Hier werden sämtliche Aktionen, die auf das Objekt angewandt werden protokolliert.
- **Agent Entity:** Diese dient der Identifikation, welche Person, Organisation oder auch Software einem Ereignis im Lebenszyklus des digitalen Objektes zuzuordnen ist.

- **Rights-Entity**: Hier wird festgehalten, ob, bzw. wer welche Erlaubnis zur Vervielfältigung des Objektes hat, bzw. auch das Gestatten der Umwandlung in andere Formate

7 Datenträgerarchivierung – die Lösung?

Zentrales Thema jeglicher Archivierung, egal ob analog oder digital, stellen die verwendeten Datenträger dar. Die Frage stellt sich hier nach dem Einsatzzweck des Datenträgers. Soll dieser dazu verwendet werden, mehr oder weniger schnellen Zugriff auf die archivierten Informationen zu ermöglichen oder soll er tatsächlich ausschließlich dazu zu dienen, den Bitstream über die Zeit zu archivieren. So dürfte deutlich werden, wenn man an Terabytes von Daten schnell herankommen möchte, diese nicht erst von einem langsamen, sequentiellen Datenträger wie Mikrofilm eingelesen und redigitalisiert werden können. Da auch CD-ROMs mit etwa 700 MB, DVDs mit 4,5-9 GB und Blu Rays mit derzeit max. um die 130 GB diese Kapazitäten nicht bieten, kommen nur schnelle Festplatten infrage. Festplatten sind allerdings keine l3-tauglichen Datenträger. Hier hat man entweder die Wahl, wie beim 3-2-1-Backup gezeigt, mehrere Datenbestände anzulegen, indem man Backups anfertigt und die Datenträger auslagert oder aber man trifft eine Auswahl, was man archiviert, konkret also: man beschränkt sich hinsichtlich der Datenmenge. Leider ist das mit Verlust an Archivgut einhergehend, wenn nicht alles archiviert werden würde.

In der Praxis wird man sich auf einen Kompromiss aus Archivierung essentieller archivierungswürdiger Daten beschränken und die Datenträger entsprechend auslagern. Archivierungswürdig sind dabei Eingangsdaten sowie zugehörige Software und Bearbeitungsvorgänge. Sind nur Endergebnisse relevant, kann man auf die Archivierung der Eingangsrohdaten verzichten. Wichtig ist in jedem Fall die Prozess-Software, die der Bearbeitung der Daten dient, entsprechend zu archivieren. Neuroth spricht in diesem Zusammenhang treffend davon, dass es DIE technische Lösung nicht gibt. Budget, Nutzungsszenarien, Art der digitalen Sammlung oder schlichtweg spezielle Rechenanlagen, die für sich Unikate darstellen, wie z.B. der CERN-Rechner für Teilchenphysik, verursachen unterschiedliche Strategien in der Archivierung. Die am CERN anfallenden Datenmengen im teilweise sogar Petabyte-Bereich können mit derzeitigen Mitteln heute noch nicht archiviert, sondern lediglich für die Nachwelt interpretierbar, aber nicht reproduzierbar bereitgehalten werden (Neuroth, 2009, S.11.1f.).

7.1 Heute verfügbare langlebige digitale Datenträger

Für rein von der Datenmenge mit geringerer Kapazität anfallendem archivierungswerten Material bestehen heute sehr wohl Lösungen in Form kostengünstiger, langlebiger Datenträger. Erwähnt wurde bereits die Mikroverfilmung. Diese mag zwar ein gangbarer Weg für die reine Archivierung sein, stellt jedoch nur eine unbefriedigende Lösung dar, wenn es ohne den Zwischenschritt einer Redigitalisierung um den Zugriff auf Datenarchive geht. Außerdem bieten nur einige Dienstleister die

Mikroverfilmung und Redigitalisierung an. Wenn DVDs oder Blu Rays ausreichende Kapazität bieten, so bietet sich heute die M-Disc als Lösung an.

Die M-Disc soll hier laut Hersteller eine tausendjährige Haltbarkeit der Datenträgerschicht bieten. Diese rührt von der Gesteinsschicht als Informationsträger her, in welche die Daten eingraviert werden. Auch wenn dem Hersteller hier kommerzielle Interessen unterstellt werden könnten, muss man zugeben, wenn man an die einleitend erwähnten Steintafeln als Zeitzeugen antiker Kulturen denkt, dass trotz fehlender Erfahrungswerte die M-Disc auf jeden Fall eine bessere Wahl ist als klassische beschreibbare DVDs oder Blu Rays, die lichtempfindlich sind. Besonders deutlich wird dies, wenn man bedenkt, dass eine M-Disc in derzeit handelsüblichen Brennern beschreiben werden kann. Mit Stand 2016 kostet somit eine Archivierung von 25 GB an Daten nur weniger als 100 €, was den Brenner und den Datenträger beinhaltet. Braucht man höhere Kapazitäten kann man eine höhere Anzahl an Datenträgern beschreiben und auch hier nach dem 3-2-1-Backup vorgehen.

Theoretisch ließe sich auch hier die Gravur mit Licht und sehr starken Vergrößerungslinsen entziffern. In der Praxis wird man allerdings, um den Datenträger in tausend Jahren lesen zu können, einen Brenner samt Computer, der ihn beschrieben hat, nachkonstruieren müssen, d.h. es werden Baupläne für dessen Rekonstruktion gebraucht. Diese müssen zugänglich gehalten werden und können daher entweder physisch auf Papier oder auf Mikrofilm den M-Discs beigelegt werden. Ein Problem besteht allerdings auch bei der LZA mittels M-Discs. Während nämlich grob bearbeitete Steine der Steinzeit mit geringen Datenkapazitäten wesentlich unempfindlicher gegen mechanische Beschädigungen waren, können M-Discs dagegen einfach zerbrochen und damit physisch vernichtet werden, was letztlich trotzdem den Verlust der Daten bedeutet.

7.2 Die GlassMasterDisc

Ähnlich wie die M-Disc verhält es sich mit der GlassMasterDisc. Diese speichert Daten nicht auf Stein, sondern in Spezialglas. Auch diese Disc soll eine Haltbarkeit von 1.000 Jahren aufweisen und temperaturunempfindlich sein. Genau wie die M-Disc lässt sie sich in handelsüblichen Brennern beschreiben. (Sylexx, 2011, S.2)

7.3 Die 5D Glass Disc

Das derzeitige Nonplusultra, entwickelt an der University of Southampton stellt die 5D Glass Disc dar. Sollte sich diese Disc durchsetzen, dürften für einige Zeit sämtliche Archivprobleme gelöst sein. Sie soll dabei unglaubliche 13.8 Mrd. Jahre (!) halten, Temperaturen von 190 Grad Celsius vertragen und Speicherplatz von 360 TB bieten, was bereits an ein kleines Rechenzentrum denken ließe (Vincent, 2016). Da es hierzu aber noch keine Erfahrung gibt, sind

diese Aussagen kritisch zu hinterfragen. Selbiges gilt für die M-Disc und GlassMasterDisc.

7.4 Lagerorte

Für LZA spielen nicht nur die Wahl von Dateiformaten, Datenträger und Archivierungsstrategien eine Rolle, auch die Wahl des Archivierungsortes gehört zu einer LZA-Strategie. Hier ist v.a. an die richtige Klimatisierung zu denken, den – soweit möglich weitgehenden – Schutz vor Naturkatastrophen, aber auch vor menschlichen Einflüssen wie Kriegen oder Diebstahl. Solche sichere Lagerorte wären in Österreich z.B. der „earthDATAsafe“ in Kapfenberg, wo man Datenträger und Services in Form von Rechenanlagen in Bergwerkstollen gegen Gebühr für Private wie Unternehmen anbietet (Kapsch, 2017). Ein anderer Ort wäre der sog. Regierungsbunker in Salzburg, wo das Bundesrechenzentrum ebenfalls tief unter der Erde Backups seiner Datenbestände aufbewahrt (ORF [2], 2016). Neben den physischen Lagerorten gibt es aber auch mehr und mehr virtuelle Lagerorte in Form von Cloud-Lösungen. Hierbei hängt es vom jeweiligen Anbieter ab, welche Services er bietet und v.a., ob die Daten in der Cloud auch verteilt in mehreren Rechenzentren gespeichert werden. Nur dann sind sie nämlich weitgehend sicher für den Fall dass ein Standort untergehen sollte. (ORF [3], 2016)

Bedenken muss man jedoch bei Datenträgern, dass man immer auf die einleitend genannten Lesegeräte angewiesen ist, was u.u. die Archivierung der Baupläne beinhaltet sowie die Datenträger immer der physischen Vernichtbarkeit unterliegen. Ersterem beugen hier i.d.R. jedoch die Hersteller selbst vor, für Letzteres ist dagegen jeder selbst verantwortlich. Datenträger sollten hier ordnungsgemäß gelagert sowie schon nach dem Beschreiben am besten in einem anderen Lesegerät testweise gelesen und noch optimaler, gleich dem erwähnten 3-2-1-Backup-Konzept folgend, mehrere Exemplare der Datenbestände archiviert werden. Einen davon sollte man verteilt außer Haus aufbewahren. Privatpersonen werden hier eher zu Cloud-Lösungen greifen oder zu M-, bzw. GlassDiscs, KMUs könnten derartige Datenträger zudem in speziellen Bunkern, wie dem „earthDATAsafe“ gegen Gebühr lagern lassen und Regierungsinstitutionen wie die Österreichische Bundesregierung können es sich auch leisten, selbst einen solchen Bunker anzulegen.

7.5 Emulation

Emulation zielt darauf ab, veraltete Systeme, die nicht mehr verfügbar sind, weil sie kaputt gingen, man sie nicht mehr herstellt, wartet oder Sicherheitslücken behebt, auf anderen Systemen zu emulieren. Daten, die an diese Systeme gekoppelt sind, weil sie ausschließlich dort funktionsfähig abrufbar sind, wären dann verloren. Hier greifen Virtualisierungstechniken, welche die ursprüngliche Umgebung abbilden und damit die weitere Nutzung der Daten ermöglichen. Genutzt wird diese Technik vielfach auch, um physische, moderne und teurer Hardware bestmöglich zu nutzen, indem z.B.

Server virtualisiert werden. Hier wird also ein aktuelles System ebenfalls in einem aktuellen System emuliert. Dadurch ergeben sich neben dem Kostenvorteil auch Vorteile hinsichtlich der sog. Live-Migration im Sinne eines Umzugs von einer emulierten Umgebung in eine andere auf eine andere physische Maschine (Larson, Carbone, 2009, S.334ff.).

Das Konzept selbst ist dabei nicht neu. Erinnerung sei auch hier an den Commodore 64. Der damals verfügbare Commodore 128 vom gleichen Hersteller verfügte „serienmäßig“ über einen Emulationsmodus, in welchem er einen vollständigen C64 emulierte – leider aber eben nicht zu 100%, da doch vereinzelt Probleme auftraten. Bezogen auf die Strategie einer LZA durch Emulation muss man sich somit darüber im Klaren sein, dass auch hier Inkompatibilitäten in Kauf zu nehmen sind. Das Problem ist aber, wenn man es erst nach dem Umstieg auf eine virtualisierte Umgebung als Spätfolge merkt, dass dort und da etwas nicht korrekt abläuft. Heutige Hypervisoren stellen hier nur emulierte Hardware, allerdings keine echten Schnittstellen und v.a. keine speziellen Eingabegeräte zur Verfügung. Zu denken wäre ganz aktuell an Registrierkassen, aber auch an ältere Scanner mit Parallelport-Schnittstellen oder Light Pens. So denke man z.B. an obigen Commodore 64. Für diesen Computer existieren Emulatoren in Hülle und Fülle. Keinem davon ist es allerdings gelungen, den Light-Pen anzusteuern. Der Grund liegt in der fehlenden physischen Schnittstelle heutiger PC-Systeme. Konsequenterweise läuft auch emulierte Zeichensoftware nicht ordnungsgemäß, die den Light-Pen benötigen würde.

Selbst wenn man umständlich einen USB-Port mit viel Bastel- und Lötarbeit kompatibel dazu machen könnte, mangelt es immer öfter an erforderlichen Kathodenstrahlbildschirmen in der heutigen Zeit. Auch Systeme, die ursprünglich nicht für die Ausführung anderer Software bestimmt waren, allerdings trotzdem dazu benutzt werden, sind ein Zeichen, wo der Emulation Grenzen gesetzt sind. Jeder der z.B. schon versucht hat, das PC-Spiel Doom auf einer Smartphone-Portierung zu spielen, weiß um die schlichte Unmöglichkeit dieses Unterfangens, denn mit den zu langsam reagierenden und kryptischen Eingabeschnittstellen kann kein einziger Spiel-Level vernünftig absolviert werden. Für den C64, Amiga, Atari ST und sämtlich damaligen Home Computer Systeme und Videospiele fehlt es heute schlichtweg an Eingabegeräten wie Joysticks und kompatiblen Mäusen. Genau dies ist auch das Problem der LZA von Computerspielen. Die logische Konsequenz: ein zuverlässig funktionierendes Archiv für Computerspiele gibt es bis dato nicht (Huth, 2009, S.17.122).

Führt man sich vor Augen, dass die C64-Zeit gerade mal 20-30 Jahre zurückliegt und man es nicht einmal schafft, derartige Systeme doch jüngeren Datums zu emulieren, dann kann man wohl kaum davon ausgehen, dass die heutigen Systeme in 1.000 Jahren noch emuliert werden können. Lange vor dem (M)icro(s)oft (D)isk (O)perating (S)ystem [MS-DOS] gab es Lochkarten. Die

Unterstützung für Lochkartenleser in MS-DOS war de facto nicht gegeben und für Bänder nur mittels sog. Gerätetreiber, der das jeweilige Gerät ansteuert. Dieses Grundprinzip: kein passender Gerätetreiber, keine Unterstützung, gilt auch heute noch sowohl für Linux und Windows. Auch von aktueller Software, die nicht länger gewartet wird, kann nicht erwartet werden, dass sie migriert wird. In diesem Zusammenhang war kürzlich auf der Webseite des Österreichischen Rundfunks zu lesen, dass sich das Web von Altlasten befreit. Gemeint war damit der Flash Player zur Wiedergabe von multimedialen Inhalten. Bezogen auf die LZA liegt hier ein konkretes und aktuelles Beispiel vor, wie Inhalte in Zukunft nicht nutzbar sein werden, sofern man keine emulierte Umgebung mit der entsprechend älteren Software laufen lässt. (ORF [1], 2016). Eine solche muss aber erst einmal geschaffen werden, ebenso wie Schnittstellen zu allenfalls alten Lochkarten- und Bandlesegeräten. Da dies faktisch nicht erfolgt, sind der Emulation letztlich Grenzen gesetzt.

Die Emulation mag somit übergangsweise ein guter Ansatz sein, möglicherweise funktioniert sie sogar auch jahrzehntelang – was aber dann? Für Hardware, die mehrere Generationen zurückliegt, darf bezweifelt werden, dass weiterhin Emulatoren für kommende Generationen geschrieben werden. Es ist auch hier einmal mehr und ausschließlich möglich, Hardware nachzukonstruieren, so wie z.B. mit einem kürzlich erschienenen C64-Joystick, der die Hardware des C64 beinhaltete (natürlich wurden die Schaltungen mit modernster Elektronik nachgebaut) und daher auch entsprechend wie das Original funktionierte. Genau deswegen setzt dies genaueste Dokumentation der Spezifikationen heutiger Systeme voraus. Sollten kommerzielle Verwertungen allerdings nicht zu erwarten sein, wäre die Entwicklung zu aufwändig bei zu geringem Nutzen. Rothenberg ist hier entschieden zu widersprechen, denn diesen Umstand bedachte er wohl nicht. Immerhin setzt er vollständig auf Emulation, was nicht nur den fehlenden Geräten und Schnittstellen älterer Systeme widerspricht, sondern auch dem OAIIS-Modell, wie noch gezeigt wird:

“Fortunately, it is not necessary to preserve physical hardware to be able to run obsolete software. Emulators—programs that mimic the behavior of hardware—can be created to take the place of obsolete hardware as needed. Assuming that future computers will be orders of magnitude more powerful than ours, future users should be able to ask their computers to generate emulators for obsolete systems on demand.” (Rothenberg [0], 1999, S.15)

Die Fiktion des Raumschiffes Enterprise mag die Problematik verdeutlichen. Sollten eines Tages tatsächlich die dortigen Replikatoren zur Verfügung stehen, die aus Molekularstrukturen sämtliche Wunschgegenstände replizieren können, würde die ersten Datenbankabfragen dieses Replikators nach einem Commodore C64 mit einem alten Spiel namens „Donkey Kong“ in etwa wie folgt ablaufen (Funk, 2009, S. 8.16ff.):

- `SELECT 'C64' FROM Computer.Bauplanarchiv`

(=die Hardware) Unmittelbar darauf könnten noch Abfragen folgen wie

- `SELECT 'C64-ROM-Listing' FROM Softwarearchiv.OS`

(=das Betriebssystem) und

- `SELECT 'Donkey Kong' FROM C64.Games WHERE Spielstand='kurz_vor_Sieg';`

(=Anwendungssoftware, in diesem Fall das Spiel mit den Daten zum Spielstand kurz vor Ende)

Hat man hier richtig archiviert, könnte Captain Picard im 24. Jahrhundert Donkey Kong spielen. Wehe aber, wenn die Konstruktionspläne des C64 nicht zur Verfügung stehen, d.h. also nicht langzeitarchiviert worden sind. In diesem Fall könnte auch der Enterprise-Computer wenig helfen. Backups auf 5 1/4-Disketten mit Donkey Kong dürften mangels der Laufwerke und verlorenen Magnetisierung einerseits sowie völlig verrosteter C64-Netzteile und geplatzter Kondensatoren auf der Hauptplatine nicht mehr zielführend sein. Bezogen auf den Enterprise-Computer möge man hier daran denken, dass obiger C64-Computer nun zwar repliziert wäre, Captain Picard allerdings nach wie vor nicht in der Lage wäre zu spielen. Dazu müsste er nämlich wissen, dass er vorher den Befehl „RUN“ eingeben müsste.

Konkret bedeutet somit Langzeitarchivierung für Emulation drei Dinge:

1. Konstruktionspläne aufzubewahren. Die dazu taugliche Möglichkeit wäre hier der bereits genannte Mikrofilm. (=Hardware)
2. Die Bitstreams der Software müssen zur Verfügung stehen, sei es ebenfalls durch Mikrofilm und Redigitalisierung, sei es durch ständig umkopierte oder durch langlebige Datenträger für die man auch Laufwerke und Lesesysteme über die Zeit hindurch produziert. (=Software)
3. Die Bedienungsanleitungen müssen vorhanden sein. (=Dokumentation)
4. Ein System, wo das Ganze ablauffähig ist. (=Host-System, bei der Enterprise wäre das das fiktionale Holodeck)

Wie Huth treffend zu dieser Problematik anführt, ist für LZA auch die Archivierung alter Bedienungsanleitungen von Hard- und Software erforderlich. Da diese früher noch gedruckt vorlagen, ist hier entweder Digitalisierung erforderlich oder die platzintensive Lagerung der physischen Druckwerke (Huth, 2009, S.8.30). Da es jedoch um LZA digitaler Daten geht, wird auch hier einmal mehr deutlich, dass das erste Kapitel zur Digitalisierung zwangsläufig in einer LZA-Strategie beherzigt werden muss. Dem steht jedoch Hughes entgegen, da Platzersparnis bei ihr keinen ausreichenden Grund für Digitalisierung darstellt. Der Sinn dahinter steckt im Fokus auf den Erhalt einzigartiger Werke denn Massenaufgaben von Büchern (Hughes, 2004, S. 51). Allerdings darf stark daran gezweifelt werden, dass heute z.B. noch sehr viele Final Cartridge III-Handbücher (ein Steckmodul des Commodore 64) in ausreichender Zahl physisch und v.a. qualitativ hochwertig erhalten sein dürften, waren die Zielgruppe doch meist Kinder, die die Schnellladefunktion nutzen wollten. Insofern muss man diese Aussage kritisch hinterfragen, besonders vor dem Hintergrund der jahrhundertelangen Erhaltung, wenn eine

Anleitung durch Kinderhände ging, die wohl kaum auf Erhalt, denn auf Spiel und kaputt machen aus sind.

7.6 Migration

Während die Emulation komplette Systeme abbilden möchte, ist es Ziel der Migration, nur die Daten zu erhalten und für künftige Systeme nutzbar aufzubereiten. Dabei ist problematisch, dass Software besser ausschließlich in emulierten Umgebungen langzeitarchiviert wäre. Dies deshalb, da Software für neue Systeme immer adaptiert werden muss aufgrund anderer Hardware und erweiterter Maschinenbefehlssätze. Die Schwierigkeit besteht hier aber darin, die bei der Emulation bereits genannte fehlende Hardware nicht zur Verfügung zu haben. Damit wäre die Neukompilierung der Software unter der neuen Umgebung erforderlich. Dies wiederum würde das Vorhandensein des Quellcodes und auch eines Compilers für das neue System voraussetzen. Da hierbei allerdings oft Altlasten mitgeschleppt werden würden, ist eine solche Vorgangsweise nicht zielführend. Erinnerung sei an Windows, das bis weit in die 32- und auch 64-Bit-Versionen noch Code aus der 16-Bit-Zeit mitschleppt, was der Kompatibilität dient. Immerhin befreite man sich bei Microsoft schon von der Unterstützung alter 32-Bit-Hilfedateien und 16-Bit-DOS-Programmen. Die Unterstützung ist zwar nach wie vor gegeben, muss allerdings explizit nachinstalliert werden.

Die Migration zerfällt dabei in zwei Teilaufgaben:

1. Die Bitfolge des jeweiligen Digitalisates zu erhalten, nicht zwingend allerdings den ursprünglichen Datenträger.
2. Die Wahrnehmung des Inhaltes muss gewahrt werden.

Während der erste Punkt recht einfach durch Umkopieren auf aktuellere Datenträger zu bewerkstelligen ist, stellt der zweite Punkt die eigentliche Herausforderung dar (Borghoff, et. al., 2005, S.489f). Dem hier bei Borghoff angedachten Museumsansatz, wo entsprechende Wiedergabeumgebungen betriebsbereit gesammelt werden, um eben die ursprüngliche Wahrnehmbarkeit zu erhalten, ist aus LZA-Sicht deutlich entgegenzutreten. Kein elektronisches System der heutigen Zeit würde hunderte von Jahren betriebsbereit überdauern. Auch hier ist die einzige Möglichkeit nicht die Geräte selbst, sondern deren Konstruktionspläne und Steuerungssoftware zu bewahren. Ohne spezielle Eingabegeräte endet nämlich wie beim erwähnten Commodore 64 die Strategie der Migration. Daran mag auch das noch zu besprechende OAIS-Modell, das ja vollständig darauf setzt, wiewohl es auch der Emulation offen ist, nichts ändern und zwar, weil die Emulation auf Dauer nicht funktioniert – die Migration ohne potentielle Rekonstruierbarkeit der Hardware allerdings leider auch nicht (Brübach, 2009, S.4.12)

7.7 Spezialfälle

Migration zielt nicht nur darauf ab, Bitstreams in andere Container zu überführen, sondern auch die Datenträger selbst zu migrieren. Dabei unterscheidet man mehrere Arten an wichtiger Strategien, die sich bei Ullrich [0] finden (2009, S.8.5f.):

7.7.1 Der Refresh

Dieser ist die simpelste Form, wobei hier lediglich bestehende Datenträger auf neue derselben Art kopiert werden. Der Vorteil besteht darin, dass es sich um das ursprüngliche Format handelt, die Daten frisch geschrieben und damit neuerlich entsprechend der jeweiligen zu erwartenden Lebensdauer des Datenträger haltbar sind. Der Nachteil ist, dass man keine Zukunftssicherheit für die LZA gewinnt, wenn die Unterstützung des Datenträgerformates ausläuft oder keine Laufwerke wie die einleitenden Band-Lesegeräte der NASA mehr verfügbar sind.

7.7.2 Die Replikation

Diese funktioniert ähnlich des Refreshs, mit der Ausnahme, dass eben kein ursprünglicher Datenträger, sondern ein anderer aktuellerer Bauart genommen wird. Der Vorteil ist, dass man sich der jeweiligen Zeit angepasst hat und mittelfristig wieder mit Versorgung durch Lesegeräte bezüglich der LZA rechnen kann, der Nachteil dagegen ist, dass Software z.B. nicht ordnungsgemäß auf neue Datenträger zugreifen kann. So kennen ältere Windows-Versionen vor Windows 95 OSR-B USB-Datenträger gar nicht sowie große Dateisysteme jenseits der 2 GB ebenfalls nicht. Hat man noch alte per DBLSPACE komprimierte Volumes, kann man sich vorstellen, dass der Zugriff leider nicht funktioniert.

7.7.3 Repackaging

Dabei werden die Container des digitalen Objektes verändert, wobei die Inhaltsdaten gleich bleiben. Für LZ-Archive ist das Repackaging ein dokumentationspflichtiger Vorgang. Vereinfacht könnte man sich vorstellen eine komprimierte ZIP-Datei zu entpacken und daraus eine weniger Speicherplatz beanspruchende 7Z-Datei zu erzeugen. Beide Dateien sind nicht verlustbehaftet, stellen aber andere Ansprüche an die jeweilige Entpack-Software. Verfügt man auf künftigen Zielplattformen nicht über einen 7Z-Entpacker wie 7zip, sondern nur über Entpacker für das nach wie vor weiter verbreitete ZIP-Format, so kann man den Content nicht mehr auspacken. Für Repackaging muss es also triftige Gründe geben, bevor man es anwendet.

7.7.4 Transformation

Während Refresh, Replikation und Repackaging noch weitgehend schonend unter Erhalt des ursprünglichen Content-Bitstreams vonstattengehen, ist die

Transformation ein schwerwiegender Eingriff in die Substanz. Der Bitstream wird dabei nämlich verändert und in ein neues Format gebracht. Notwendig kann ein solches Vorgehen dann werden, wenn das Ursprungsformat generell keine Unterstützung mehr in künftigen Systemen erfährt, d.h. weder Emulatoren zur Verfügung stehen, nicht mit Entwicklung solcher oder der Rekonstruktion der ursprünglichen Hardware zu rechnen ist und keine Möglichkeiten der „Nach-Erzeugung“ des ursprünglichen Contents mit vertretbarem Aufwand bestehen. Damit ist gemeint durch das Ursprungsprogramm den Content gleich in einem verbreiteten Format zu speichern. Zu denken wäre hier z.B. an alte Ami Pro-Textdokumente im „.sam“-Format, die man auch im Word 2.0-Format speichern hätte können, womit sie sich wesentlich leichter migrieren ließen, als mit Konvertierungsfiltren für das ältere MS-Office 4.3. Mit Layoutproblemen und nicht hundertprozentiger Unterstützung für alle Features beim Migrationsprozess von anderen Dateiformaten ist auf alle Fälle zu rechnen. Ähnlich verhält es sich bei Video und Musik-Daten.

8 Langzeitarchivierung nach OAIS

Der Open Archival Information Standard ist heute de facto die Referenz der digitalen Langzeitarchivierung. Bemerkenswert ist, dass OAIS nach wie vor nur ein Modell ist und keinerlei normativen Charakter hat. OAIS geht dabei ausschließlich von schon digitalisierten Daten aus. Anzumerken ist allerdings, dass ohne entsprechende Hardware und Dokumentation auch OAIS-basierte Archivierung zum Scheitern verurteilt wäre:

„Die Gedächtnisorganisationen werden in Zukunft eben auch Papier enthalten müssen (Anm.: im Sinne von Lagerung), es treten neue Aufzeichnungsformen hinzu, die die alten keineswegs vollständig verdrängen werden. Ebenso wie sich das noch vor wenigen Jahren propagierte "papierlose Büro" als Hirngespinnst erwiesen hat und, viel bescheidener, heute nur noch vom "papierarmen Büro" gesprochen wird, sind Überlegungen zu einem vollständigen Medienbruch bei der Archivierung realitätsfremd. Das OAIS berücksichtigt Bestehendes: Es ist gerade deshalb ein Modellansatz und Standard, der damit auch Einfluss auf zukünftige Arbeitsmethoden im Archiv nehmen wird.“ (Brübach, 2009, S.4.6.f.)

Damit ist das grundlegende Problem umrissen, sich hier nicht von technischen Archivierungssystemen abhängig zu machen. Der beste Datenträger, der tausende oder Mrd. von Jahren überdauert, wie die M-Disc, GlassMasterDisc oder 5D Glass Disc nutzt nämlich nichts ohne Lesegeräte, die Kenntnis der Dateiformate sowie entsprechender Ausgabegeräte. Dementsprechend ist der Standard offen gehalten. Kritisch muss man hinterfragen, dass er möglicherweise zu offen gehalten ist. Vertrauenswürdige digitale LZ-Archive nach DIN 31644 sind nämlich das Ergebnis eines mehr als 20 Jahre andauernden Fachdiskussionsprozesses. Maßgebliche Stakeholder waren die australische Nationalbibliothek, die DRAMBORA-Gruppe für die EU, Trac und die NASA in den USA, bzw. die nestor-Gruppe in Deutschland. Alle diese Gruppen hatten unterschiedliche Vorstellungen. Die ursprüngliche Definition von Langzeitarchivierung war hierbei:

„Organisation (bestehend aus Personen und technischen Systemen), die die Verantwortung für den Langzeiterhalt und die Langzeitverfügbarkeit von Information in digitaler Form sowie die Bereitstellung für eine bestimmte Zielgruppe übernommen hat.“ (Keitel, Schoger, 2013, S.5)

In der Diskussion zur Erweiterung dieses klassischen Ansatzes einer Archivdefinition trat auch der Vorschlag hervor, digitale Objekte unmittelbar bei ihren Entwicklern archivieren zu lassen. V.a. in Hinblick auf deren allumfassendes Wissen diesbezüglich erschien diese Strategie zwar logisch, wurde jedoch schnell wieder verworfen. Vielfach waren die Hersteller einfach nicht bereit dazu, alte und nicht mehr zeitgemäße digitale Objekte zu verwalten, was zeit- und personalintensiv wäre. Auch allfällige Supportkosten für alte Objekte fallen hier ins Gewicht. Zum Spaß gibt es keinen End of Life-Cycle von Herstellern digitaler Objekte, wobei v.a. an Software zu denken ist. Daher mussten letztlich doch die sog. Gedächtnisorganisationen wie eben Bibliotheken, spezielle Bundes- und Nationalarchive, Universitäten oder

Museen dafür einstehen. Weitere Schwierigkeiten bereiteten erste Entwürfe, die noch keine Trennung von Daten und Information vollzogen. Konkret waren daher die Abstraktionsmodelle noch nicht ausgereift, lizenztauglich zu sein. Auch Bedenken hinsichtlich der Sicherheit vor unbefugtem Zugriff und Bearbeitungen spielten eine Rolle. Hier wäre an die Auslagerung in Cloud-Systeme zu denken. (Keitel, Schoger, 2013, S.6ff.)

Kritik am ersten OAIS-Standard von 2003 ließ somit nicht lange auf sich warten. Wie immer, wenn mehrere Akteure beteiligt sind, kommt es zu Streitigkeiten über Ziele und Inhalte. So wirkten in der Arbeitsgruppe ab dem Jahr 2004 Vertreter der Gedächtnisorganisationen, Nutzer, Hersteller von Content und Software sowie Archivare mit. Der Wunsch nach einem allgemeinen Kriterienkatalog, der für alle gilt, wurde dabei schnell laut (Keitel, Schoger, 2013, S.33). Diesem Wunsch wurde mit dem nestor-Memorandum von 2006 letztlich Rechnung getragen. Die wichtigsten nestor-Kriterien bezogen auf DIN31644 sind dabei, was den Umgang mit digitalen Objekten im Allgemeinen anbelangt:

- Integrität der Objekte
- Authentizität der Objekte
- Verfügbarkeit der Objekte
- Vertraulichkeit der Objekte

wobei sämtliche dieser Aspekte der IT-Sicherheit zugeordnet werden können, während zum Zweiten die Langfristigkeit durch folgende Aspekte abgedeckt wird:

- langfristige Auffindbarkeit und Referenzierbarkeit der Objekte
- langfristige Interpretierbarkeit der Objekte

Sämtliche obigen Punkte verstehen sich dabei inklusive dem Erhalt ihrer Metadaten. (nestor, 2006, S.15) Der Kriterienkatalog wurde dabei von einigen der im Quellenverzeichnis genannten Personen mitverfasst, wobei DIN 31644 starken Bezug auf OAIS von 2003 nimmt, das 2012 neuerlich als ISO 14721 verabschiedet wurde. Federführend waren auch das britische Digital Curation Centre und die Digital Preservation Coalition (Oßwald, 2009, S.4.2). Im Folgenden werden die einzelnen Kriterien detailliert beleuchtet:

Integrität der Objekte: Hier wird auf den vollständigen Erhalt der Daten gesetzt. Es darf also nichts bewusst, unbewusst oder durch technisches Versagen verloren gegangen sein. Ist dem so, so ist das Objekt umgehend zu ersetzen mittels Restore aus einem unbeschädigten Bereich.

Authentizität der Objekte: Darunter versteht man den Erhalt der signifikanten Eigenschaften eines digitalen Objektes, die unverändert bleiben müssen, um

das Objekt als solches zu identifizieren. Abgewandelt und dokumentiert werden müssen dabei alle sonstigen Änderungen, wie z.B. Wechsel des Dateityps im Sinne einer verlustfreien Migration. Unzulässig wäre hier z.B. die Migration in ein verlustbehaftetes Format.

Verfügbarkeit der Objekte: die Verfügbarhaltung stellt einen zentralen Aufgabenbereich des Archives dar. Hier geht es um die eigentliche Archivarbeit wie das rechtzeitige Migrieren, Testen und nach 3-2-1-Backupstrategie für redundante Vorratshaltung der digitalen Objekte zu sorgen.

Vertraulichkeit der Objekte: Hier gilt es abzuwägen, ob Zugriffe und Sicherheitsmaßnahmen auf digitale Objekte anzuwenden sind. Das Archiv hat hier unter Umständen auch den Zugang zu archivierten Objekten zu beschränken. Zu denken wäre hier z.B. an Archive zur nationalen Sicherheit mit militärischen Geheimnissen wie eben der erwähnte sog. Regierungsbunker. Erst wenn diese Vertrauenswürdigkeit gegeben ist, ist auch die ISO-Norm 16363 erfüllt.

langfristige Auffindbarkeit und Referenzierbarkeit der Objekte: Hier geht es um die Pflege der Metadaten und die Organisation eines DMS, das es ermöglicht in angemessener Zeit entsprechende digitale Objekte analog zu klassischen Bibliotheken auszuheben.

langfristige Interpretierbarkeit der Objekte: Sowohl Inhalts- wie Metadaten müssen durch die Zielgruppen uneingeschränkt nutzbar bleiben. Um dies sicherzustellen, ist hier die regelmäßige Überprüfbarkeit der Interpretation eine Aufgabe des Archives. Bei auftretenden Mängeln ist umgehend aus redundanten Beständen der Ursprungszustand wiederherzustellen, um die Interpretierbarkeit wiederum zu gewährleisten.

Sämtliche Kriterien sind dabei auf alle Stufen der Archivierung (Aufnahme, Erhaltung und Nutzung) anzuwenden (Keitel, Schoger, 2013, S. 76ff.). Zudem gilt es, für jedes Archiv die entsprechenden räumlichen und klimatischen Bedingungen wie Luftfeuchtigkeit und Temperatur einzuhalten, für Hintanhaltung von Schimmelpilzbefall zu sorgen sowie entsprechende Beleuchtung, die Datenträgern schaden könnte auf ein Minimum zu reduzieren, was im Normalfall nur bei der Aushebung der Fall ist. (Hofmann, Wieser, 2015, S.357ff.)

8.1 Das OAIS-Modell

Konkret bildet OAIS den Archivierungsprozess in mehreren Etappen wie im Folgenden dargestellt ab:

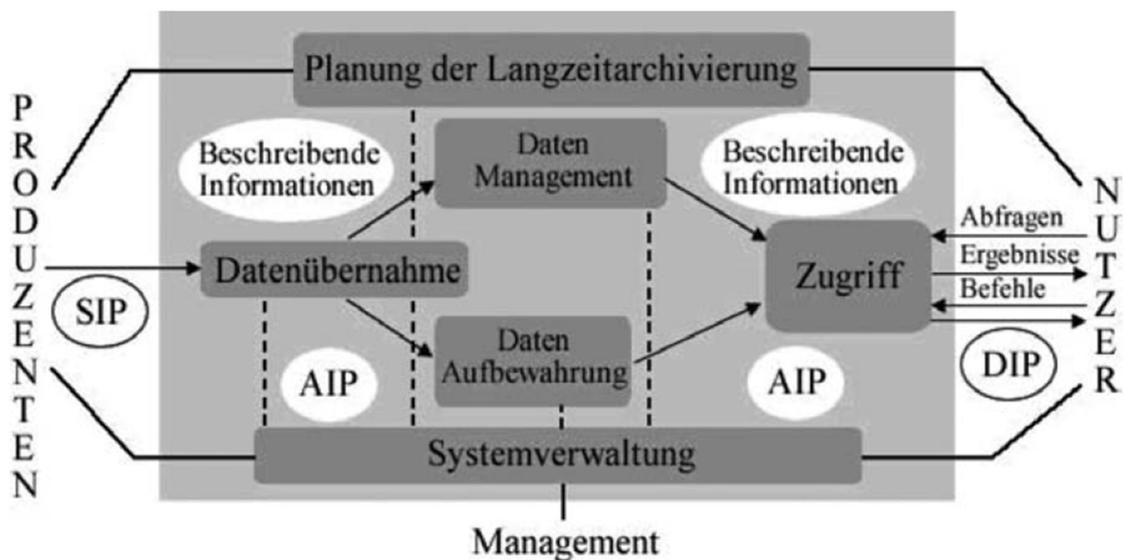


Abbildung 6: Das Funktionsmodell des OAIS (Brübach, 2009, S.4.10)

Im Wesentlichen geht es um die grau schattierten Bereiche als die eigentlichen OAIS-Aufgaben. Zu nennen wären hier (Brübach, 2009, S.4.10ff.):

8.1.1 Die Datenübernahme (Ingest)

Im Bereich Ingest ist die Vorbereitung der Übernahme entscheidend, d.h. konkret die Produzenten von Content müssen allenfalls auch die Mittel dazu bereitstellen. So wäre es z.B. denkbar, dass mit Kopierschutzmechanismen versehener digitaler Content mit einem Generalschlüssel ausgeliefert wird, damit das Archiv die Aufgaben bewältigen kann. Qualitätssicherung, Prüfung der Lesbarkeit, Verständlichkeit sowie die Herstellung der Informationspakete (AIP) in Übereinstimmung mit den Formaten und Standards des Archivs gehören hier ebenso dazu. Wenn dabei eine Vielzahl von unterschiedlichen Formaten übernommen werden muss, gehört dazu auch die Erstellung der Metainformation. Auch regelmäßige Updates des Archivspeichers sind hier zu beachten.

8.1.2 Die Datenaufbewahrung (Archival Storage)

„Achival Storage“ umfasst neben der internen Organisation, die AIPs vom Übernahmebereich in Empfang zu nehmen und einzulagern sowie für die Wiederauffindbarkeit zu sorgen. Obiges genanntes Refreshing, die Replikation, Migration und allenfalls Transformation ist ebenso Teil hiervon. Ebenso findet einmal mehr das 3-2-1-Backup sinnngemäße Erwähnung indem OAIS ausdrücklich den Vorteil redundanter Speicherung hervorhebt.

8.1.3 Datenmanagement

Datenmanagement beschäftigt sich hauptsächlich mit der Wartung der Metadaten und Verzeichnisse sowie der Verwaltung unterschiedlicher Archivdatenbanken. Wie bereits in dieser Arbeit gezeigt wurde, stellt die Erzeugung einer Datenbank anhand von Metadaten einen individuellen

Prozess des jeweiligen Archives dar. Daher können auch unterschiedliche Datenbank-Management-Systeme und DMS zum Einsatz kommen. Der Abstraktion der digitalen Objekte selbst tut dies keinen Abbruch.

8.1.4 Systemverwaltung

Hier geht es um die Beziehungen der Archivare und Nutzer als Schnittstellenfunktion zum Archiv selbst im Sinne der Betriebshard- und software. Ganz im Sinne eines Prozess-Managements gibt es Regelungen für Zuständigkeiten und Dokumentation der Arbeitsvorgänge. Rechtliche Aspekte wie Verträge zur Übergabe und Nutzung von Content sind ebenfalls Teil der Systemverwaltung.

8.1.5 Planung der Langzeitarchivierung (Preservation Planning)

Planung deckt den Bereich der langfristigen Strategie ab und ist daher zukunftsorientiert. Maßgebliche Bereiche sind Empfehlungen, in welchen Zeitabständen Updates durchzuführen sind, Migrationen erfolgen oder Überwachungen des Archives selbst vorgenommen werden. Auch die Zielgruppen, also Nutzer des Archives, können sich hier ändern, weshalb auch darauf reagiert werden muss.

8.1.6 Zugriff (Access)

Dieser Teil handelt vom Zugänglichmachen, dem Support des Archives und den Support für Nutzer, aber allenfalls auch vom Ausschluss selbiger.

Die Bezeichnungen:

- SIP meinen damit Übergabepakete (Submission Information Packages)
- AIP Archivpakete (Archival Information Packages, AIPs) und
- DIP Nutzungspakete (Dissemination Information Packages, DIPs)

8.2 Kritik an OAIS

Das grundsätzliche Problem der digitalen Archivierung ist die Adressierung von Objekten. OAIS stellt hier lediglich ein Framework dar, wie welche Prozesse zur LZA zu gestalten sind. Die eigentliche Implementierung für das jeweilige Archiv bleibt diesem ja selbst überlassen. Gefragt sind daher einheitliche Container für digitale Objekte. Erste Ansätze dazu wurden mit dem Universal Preservation Format (UPF) Mitte der 90er-Jahre des 20. Jahrhunderts geschaffen. Die darin geschaffenen Strukturen mündeten letztlich aber infolge unterschiedlichen digitalen Contents in diverse andere Standards wie eben OpenDoc, XML, das als Container für Metadaten dient, sowie in TIFF und letztlich PDF/a (Lazinger, 2001, S.189ff.). Damit ist aber einmal mehr bewiesen, dass einheitliche Standards auch unter OAIS Mangelware sind.

Der kritischste Prozess in der Archivierungsstrategie stellt dabei die Datenübernahme dar. Wie in den vorangegangenen Kapiteln erläutert wurde, gibt es eine Vielzahl digitaler Daten, die unterschiedlich gehandhabt werden.

Für OAIS stellt sich dabei die Frage, wie man sicher sein kann alles auch richtig zu archivieren. Hier ist an Regionalcodes bei DVDs zu denken, aber auch an Software, die lediglich in einer native-Umgebung ablaufen würde, da sie nur im Zuge einer Sektor für Sektor-Sicherung lauffähig wäre, und auch dann nur bei der entsprechenden Hardware, an die sie per Hash-Wert während der Installation gekoppelt wurde. Hier gab es z.B. die Unsitte einiger Programme, noch Kennungen in reservierten Bereichen von mechanischen Festplatten auf der Spur 0 zu hinterlassen. Wenn sie im Zuge eines reinen Filesystem-orientierten Kopierens vervielfältigt wurde, verweigerte die Software den Dienst (PC-Welt, 2001). Problematisch erscheint in diesem Zusammenhang, wenn man nicht merkt, funktionsunfähige Archive zu erstellen – ein Verify ergibt hier nämlich keinerlei Anhaltspunkt, dass der digitale Bitstream fehlerhaft ist. Ebenso bringt es nichts, die Setup-Dateien zu archivieren, wenn keine mechanischen Festplatten mit Spur 0 mehr existieren, bzw. künftige Dateisysteme und Betriebssysteme anders aufgebaut sind. Verpflichtende Ablieferungen von Generalschlüsseln seitens der Hersteller bringen zwecks Archivierbarkeit hier nur solange etwas, solange diese nicht insolvent werden, was natürlich auch schon vorgekommen ist. Dies ist auch der Grund für Ablieferungsverordnungen von Belegexemplaren an die Gedächtnisorganisationen. Das Problem der fehlenden Hardware, Stichwort: Spur 0-Problem, bleibt allerdings.

Was hier deutlich wird, ist, dass auch unter OAIS eine Vielzahl an ungelösten Fragen existiert. Weitere Schwächen von OAIS sind die Veränderbarkeit von OAIS selbst. So gab es 2003 eine Veröffentlichung und 2012. Wenn das Modell somit selbst einer derart schnellen Wandelbarkeit unterliegt, gewinnt man hier den Eindruck, in tausend Jahren kaum zuverlässig auf die danach archivierten Daten zugreifen zu können. Ein Mehrgewinn an LZA als wie wenn man ihn mit klassischen Methoden des schlichten Umkopierens von Daten auf neue Datenträger erreichen würde, ist hier nicht erreicht. Auch steht zu befürchten, dass ältere Standards ungültig werden und damit erst recht wieder zu Migration digitaler Daten zwingen. Da aufgezeigt wurde, dass die Migration nicht immer erfolgsversprechend ist, ist auch OAIS unter diesem Gesichtspunkt nicht besser, als obige Kopierstrategien.

Zum anderen schwächelt OAIS am Zugriffsschutz, d.h. am geschützte Netzwerk und dessen Überprüfung sowie der ausschließlichen Fokussierung auf Migration, selbst wenn es anderen Erhaltungsstrategien im Einzelfall gegenüber ebenfalls offen ist (Oßwald, 2009, S.4.12), da hier sämtlichen Möglichkeiten des mutwilligen Zerstörens sowie technischem und menschlichem Versagen Tür und Tor geöffnet ist. Zwar setzt man bei der Deutschen Nationalbibliothek auf die „wichtigste“ Strategie von OAIS, der Migration, dennoch ist man dort schon offen gegenüber anderen Erhaltungsstrategien, sollten diese sich als geeigneter erweisen, wenn die ursprünglichen Systeme veralten. V.a. die Beobachtung der aktuellen

Marktsituation ist hier Mittel zum Zweck (Brodersen, Schrimp, 2009, S8.10). Rothenberg vertritt hier den gegenteiligen Standpunkt und favorisiert die Emulation, da dabei das Original erhalten bleibt. Schlägt nämlich bei der Migration, die im Schnitt alle fünf Jahre stattfindet, eine der Zwischenmigrationsstufen fehl, möglicherweise, ohne es zu merken, wäre ein Migrationsschritt, ausgehend vom ursprünglichen Original nicht mehr möglich (Rothenberg [2], 2000, S.32ff.)

OAIS wird zwar heute von unterschiedlichen Institutionen weltweit akzeptiert. Vieles ist allerdings nach wie vor strittig. So wurden zwischen 2002 und 2007 international gleich drei verschiedene Kriterienkataloge ausgearbeitet. Problematisch dabei war, dass die einzelnen Kriterienkataloge hier mehr oder weniger nur die Anforderungen der jeweiligen Arbeitsgemeinschaften DRAMBORA, TRAC und nestor widerspiegelten. Erst später mündeten die Ergebnisse in die weniger detaillierte ISO-Norm 16343 und letztlich den nestor-Kriterienkatalog mit DIN 31644. Allerdings führten obige Streitereien und individuelle Anforderungen dazu, dass man sich im Jahre 2007 in Chicago nur auf zehn gemeinsame Grundprinzipien verständigen konnte, die tatsächlich allen digitalen LZ-Archiven gemein sein sollten. Diese wären (Keitel, Schoger, 2013, S.29):

1. Übernahme der Verantwortung zur dauerhaften Erhaltung und Pflege der digitalen Objekte für die identifizierten Zielgruppen.
2. Beleglegungspflicht über die Beständigkeit, Finanzen, Personalausstattung und Prozesse).
3. Das LZA verfügt dauerhaft über die erforderlichen Rechte (per Vertrag oder Gesetz), um die Aufgaben zu erfüllen, also konkret DRM-Umgehung, bzw. freie Belegexemplare.
4. Effektives und effizientes Geflecht einer Policy.
5. Das LZA erwirbt und übernimmt digitale Objekte aufgrund definierter Kriterien, Verpflichtungen und Fähigkeiten.
6. Integrität, Authentizität und Nutzbarkeit der dauerhaft aufbewahrten digitalen Objekte wird sichergestellt.
7. Das LZA dokumentiert sämtliche Maßnahmen, die während des gesamten Lebenszyklus auf digitale Objekte angewendet werden durch entsprechende Metadaten.
8. Das LZA übernimmt die Bereitstellung der digitalen Objekte.
9. Das LZA verfolgt eine Strategie zur Planung und Durchführung von LZA-Maßnahmen
10. Vorhandensein einer angemessenen technischen Infrastruktur zur dauerhaften Erhaltung und Sicherung der digitalen Objekte.

Bei genauerer Betrachtung sind die Punkte eins bis drei sowie acht bis zehn selbstverständlich, da ansonsten die Sinnhaftigkeit und der Betrieb eines LZ-Archives nicht möglich wäre. Bleiben effektiv die Punkte vier bis sieben. Vier und Fünf stellen dabei die besagten subjektiven Entscheidungen dar und damit die individuellen Interessen der beteiligten Stakeholder. Punkt sieben ist im eigenen Interesse des LZ-Archives zu erfüllen, da es die Bewältigung der Aufgaben erleichtert. Man muss daher den Eindruck gewinnen, als handle es sich bei diesen Punkten um Alibi-Grundprinzipien, da als tatsächlich kleinste gemeinsame Nenner ausschließlich die obige Integrität, Authentizität und die Nutzbarkeit aus Punkt 6 übrig bleiben. Gesagt werden muss fairerweise dazu aber, dass diese Prinzipien aufgrund des Faktums der generellen Knappheit von Ressourcen aufgestellt wurden. Würde ein Archiv über unbeschränkten Speicher, unbeschränkte finanzielle Mittel und eine unerschöpfliche Anzahl an Personal verfügen, könnte es auch unbeschränkt digitale Objekte archivieren.

Was Personal anbelangt, spielt dies eine wichtige Rolle bei einem weiteren Kernpunkt von OAIS: den signifikanten Eigenschaften, die ein digitales Objekt ausmachen und zwingend keiner Änderung unterworfen werden dürfen. Andere Eigenschaften sollten zwar nach Möglichkeit auch nicht angetastet werden, können allerdings, wenn es Migrationsprozesse erforderlich machen, geändert werden. Bei der Übernahme in das Archiv haben daher solche signifikanten Eigenschaften identifiziert und dokumentiert zu werden wie dies in der DIN-Norm 31644:2014-04 in Abschnitt K13 geregelt ist. Konkrete Spezifikationen werden allerdings nicht genannt, da OAIS eben nur ein Modell mit Richtliniencharakter ist. Spezifikationen könnten auch keine gegeben werden, da sich Technologien ändern, ebenso wie sich Spezifikationen selbst ändern können. Das Ziel liegt daher darin, einmal definierte Kriterien durchgehend für den Erhalt des digitalen Objektes beizubehalten:

„Die Entscheidung, welche Eigenschaften signifikant sind und welche nicht, ist eine Ermessensentscheidung, die das digitale Langzeitarchiv unter Berücksichtigung seiner Zielgruppen und seiner ökonomischen Möglichkeiten treffen muss. So ist zum Beispiel die Eignung der Erhaltungsmaßnahme, ein im Word-Format vorliegendes Dokument in ein PDF/A umzuwandeln, davon abhängig, ob die Eigenschaft <<Bearbeitbarkeit>> als signifikant betrachtet wird. Diese Einschätzung wiederum ist davon abhängig, ob die zu erwartenden Nutzergruppen des digitalen Langzeitarchivs das Dokument voraussichtlich nur lesen oder auch nachbearbeiten möchten. Nicht ohne Bedeutung ist zudem die Einschätzung des Langzeitarchivs, ob ein dauerhafter Erhalt der Eigenschaft <<Bearbeitbarkeit>> mit den aktuellen technischen Möglichkeiten überhaupt möglich und, wenn ja, bezahlbar ist.“ (Keitel, Schoger, 2013, S.56).

Diese Aussage spiegelt den Kernpunkt und die Problematiken nicht nur von OAIS sondern der gesamten LZA wider. Ableiten lässt sich daraus nämlich Folgendes:

1. Archivierung ist immer eine subjektive Ermessensentscheidung.
2. Die Kriterienauswahl von Archiv zu Archiv wird infolge Punkt 1 nicht einheitlich gehandhabt, sondern folgt nur einheitlichen Modellansätzen, nämlich der „Auswahlentscheidung“.
3. Eine Einschränkung auf Nutzergruppen erfolgt, obwohl künftige Generationen ein disperses Publikum sind, mit Ausnahme, dass sich das Archiv selbst auf einzelne spezielle digitale Objekte beschränkt (Filmarchiv, Software-Archiv, Zeitungsarchiv,...).
4. Die Qualität der Archivierung ist von finanziellen Mitteln abhängig. Damit sind menschliche Kulturen aus weniger begüterten Gesellschaften im Nachteil, wenn diese ihre Archive – wenn überhaupt – nur unzureichend pflegen und schützen können.

8.3 Die Besonderheit von OAIS

Die eigentliche Besonderheit von OAIS ist, dass es sämtlichen Kritiken zum Trotz sehr wohl Normen und Standards befolgt, die für sich gesehen in Summe eine Archivierungsstrategie darstellen. Grundsätzlich kann sich dabei jeder ein digitales Archiv nach dieser Vorgehensweise aufbauen, welches auch mehr oder weniger ausfallsicher sein mag. Das Kriterium des „Trusted Digital Repository“ erfüllt ein solches Archiv jedoch nicht. Dazu ist es zwingend erforderlich OAIS-konform zu sein und die entsprechenden Normen einzuhalten (Egger, 2005, S.18). Die spannende Frage ist aber, wenn man Archive auch ohne OAIS-konform zu sein aufbauen kann, welchen Vorteil man hier hätte. Immerhin ist die Einhaltung von Regeln, Normen und Standards de facto immer mit erhöhten Kosten und verdünnter Willensfreiheit verbunden.

Es ist daher zu klären, warum Vertrauenswürdigkeit gegeben sein muss und was darunter zu verstehen ist. Der Grund für Vertrauenswürdigkeit eines LZ-Archives liegt nun darin begründet, dass kommende Generationen nicht wissen können, welches Objekt als glaubwürdig erachtet werden kann und welches nicht. Der Ursprung der Idee des Ausschlusses verfälschender Eigenschaften an Dokumenten liegt bereits im 17. Jahrhundert, wo sich die Kirchenhistoriker Daniel von Papenbrock und Jean Mabillon mit diesem Problem auseinandersetzten (Keitel, Schoger, 2013, S.20). Verfälschende Eigenschaften können dabei bezogen auf digitale Objekte nicht nur unbefugte Änderungen an ihnen sein, sondern auch solche, die aufgrund der zu Beginn dieser Arbeit erwähnten Bitfäule entstehen. Erinnerung sei auch an die einleitend erwähnten medizinischen Daten, wenn falsche Dosierungen von Medikamenten eingelesen und fortan verabreicht werden würden.

Egal welche Institution hier mit der Archivierung befasst ist, sie hat sich zwangsläufig auch mit den rechtlichen Problematiken bei der Übernahme zu beschäftigen (Urheberrecht, Digital Rights Management). Finanzierung, die

Möglichkeit schnell auf Veränderungen in technischen, organisatorischen oder rechtlichen Bereichen zu reagieren sowie die interne Organisation und Qualitätsmanagement gehören hier ebenfalls zu den Kernaufgaben eines digitalen Langzeitarchives (Vlaeminck, 2009, S.15.4).

9 Ergebnis und Interpretation

Bei der LZA gibt es viele offene Probleme bezüglich der Art der zu archivierenden Daten. Liegen diese noch analog vor, müssen sie erst in einen Bitstream umgewandelt, also digitalisiert werden. Systembedingt ist dabei immer mit Qualitätsverlust vom Original zu rechnen. Andererseits können Daten schon digital produziert werden, wobei dann zwar kein Qualitätsverlust vorhanden ist, allerdings schleichend Bitfäule eintreten kann und jeglicher Bitstreams ohne Datenträger der Flüchtigkeit unterliegt. Die Abhängigkeit von ausschließlich maschinenlesbaren Datenträgern birgt das Problem des Verlustes einzelner Bits und damit letztlich von Information, die sich erst aus korrekter Interpretation des gesamten Streams ergibt. Fehlen wichtige Kenndaten, d.h. also Informationen, in welchem Dateiformat ein Bitstream vorliegt, ist es schwierig bis unmöglich, daraus konkretes Wissen über den Inhalt abzuleiten. Einzelne Bytes können dabei von Zeichen über Maschinenanweisungen bis zu Zahlen darstellen, je nach Breite (Byte, Word, Double Word) auch entsprechende UNICODE-Zeichen oder höhere Wertbereiche von Zahlen. Aufgabe der LZA ist es daher, den unverfälschten Bitstream vollständig zu erhalten.

Digitale LZA bedeutet somit, dass Gedächtnisorganisationen wie Bibliotheken, Museen, Rundfunk- und Zeitungsarchive mehr und mehr sog. Digitalisate zu verwalten haben. Es kann sich dabei auch um Objekte von Herstellern handeln, die nicht mehr weiter gewartet werden, da sie nur Kosten verursachen würden (End of Life), wenn man sie dort selbst weiter verfügbar halten würde. Der Erhalt liegt andererseits in der menschlichen Kultur begründet und v.a. im Zugriff auf damit bearbeitete oder erstellte andere Objekte, wie z.B. durch Software erzeugte Texte oder Grafiken. Ohne LZ-Archivierung der technischen Abspiel- oder Verarbeitungsgeräte ist der Auftrag jedoch nur unzureichend erfüllt. Die Lesesysteme altern ebenfalls und werden i.d.R. nur für eine begrenzte Zeitepoche gebaut. Somit hat ein Archiv auch die Aufgabe neben dem Bitstream für ablauffähige Umgebungen, also Hardware zu sorgen, was allerdings nicht praktikabel ist für hunderte und tausende von Jahren.

Zu nennen wären Refresh, die Replikation, solange möglich, Repackaging und Transformation. Datenträger werden allerdings auch weiterhin benötigt. Um dem Datenverlust in Zukunft vorzubeugen, ist das 3-2-1-Backup von Bedeutung, wo man über mind. dreifache Datenbestände auf mind. zwei verschiedenen Datenträgersystemen verfügt und einen Datenträgersatz davon auslagert. Egal wie man dies organisiert, es werden auch Strukturen benötigt, um darauf zuzugreifen. Hier finden sich Dateisysteme wie -formate, wobei deren Vielfalt eine große Herausforderung für die LZA darstellt. Langzeitstabile Dateiformate gibt es dagegen nur sehr wenige wie PDF/a, TIFF, JPEG und das weit verbreitete PNG für statische Informationen wie Text und Bild. Allerdings muss man auch zum Schluss kommen, dass es am Anreiz fehlt, in einigen Bereichen weitere Entwicklungen zu betreiben:

„Das PNG-Format ist ein gutes Beispiel für eine praktisch optimale, kaum nicht zu verbessernde Lösung für ein wohl verstandenes und verbreitetes Problem. Es ist heute einfach nicht mehr interessant, dasselbe Problem noch besser zu lösen, da die einzige Optimierungsmöglichkeit eine gegenüber PNG noch kleinere Dateigröße wäre, aber der Bedarf daran nimmt mit der Verbreitung immer größerer Speicher- und Übertragungskapazitäten ständig ab.“ (Bárány, 2006, S.104)

Für Video existiert zwar MPEG, allerdings räumt man in Behördenkreisen wie dem Landesarchiv Nordrhein Westfalen ein, über keine konkreten Strategien zum Erhalt von Videos zu verfügen. Selbiges gilt für Computerspiele, wie Huth aufzeigte (Huth, 2009, S.17.122). Big Data Systeme der Internet-Größen (Facebook, Google), setzen hier zumindest auf verteilte Rechenzentren. Sinn macht dies v.a. deswegen, da man die Datenmengen ansonsten nicht beherrschen würde, auch weil sie häufigen Änderungen unterworfen sind

LZ-Archive dagegen haben die Kernaufgabe weniger starken Änderungen unterworfen Information über die Zeit zu retten. Wenn Änderungen erforderlich sind, haben diese nachvollziehbar protokolliert zu werden. Wesentliche Bestandteile einer LZ-Archivierungsstrategie sind daher neben den quelloffenen Dateiformaten standardisierte Prozesse, die genau diese Aufgaben regeln. Dazu bedarf es entsprechender Normen und De-facto-Standards wie dem OAIS-Modell. Zentrale Bereiche sind dabei die Übernahme, Aufbewahrung, Systemverwaltung, Management, Zugriff sowie Planung der LZA. Sichergestellt werden muss die Integrität, Authentizität, Verfügbarkeit sowie Vertraulichkeit der Objekte.

Großer Wert bei der Archivierung wird zwecks langfristiger Auffindbarkeit auf Metadaten gelegt. Diese sind vergleichbar mit einem Inhalts- bzw. Schlagwortverzeichnis. Metadaten werden heute meist automatisch erstellt, kritisch hinterfragen muss man allerdings hier, dass die Strukturen nicht atomar gehalten sind. So bietet PDF/a ein Feld für den Autor, aber keine Standards, ob zuerst der Vorname oder Nachname geschrieben werden sollte. Unterschiedliche Archivsysteme können zudem unterschiedliche Metadaten erzeugen, was dazu führen kann, dass man digitale Objekte erst recht nicht wieder findet, wenn diese doppelt oder mit falschen Schlagworten angelegt wurden. Auch sprachliche Differenzen in der Pflege von Metadaten ist einer LZA-Strategie nicht förderlich. Moderne Bibliotheken, wie die US Library of Congress sind hier federführend, was z.B. den Dublin Core Standard dafür angeht. PREMIS dagegen bietet auch faktisch Logdateifunktion für Archivzwecke.

Was Datenträger anbelangt, so bieten sich die M-Discs und die GlassMasterDisc für LZA an. Allerdings sind diese lediglich für kleinere Datenbestände ein gangbarer Weg, während die 5D Glass Disc mit 360 TB sehr wohl als archivtauglich anzusehen ist. Leider fehlen bei allen Datenträgern Erfahrungswerte. Zwar definiert die ISO/IEC 10995:2008-Norm verschiedene Testszenarien, die von extremen Temperatur bis Luftfeuchtigkeitsbedingungen

reichen, allerdings sind dies ebenfalls künstliche Bedingungen. Bewährt hat sich dagegen der Mikrofilm, für den bereits Erfahrungswerte vorliegen. Zwar ist auch hier die Lebensdauer von 500 Jahren nur geschätzt, doch benötigt man für dessen Nutzung außer Licht und Vergrößerungsoptiken nichts Spezielles, um ihn zu lesen (Keitel, 2009, S.8.32f).

Grundsätzlich konnte festgestellt werden, dass zwar alle Arten von Bitstreams ausgedruckt zu Papier gebracht werden können und damit potentiell auch audiovisuelle Daten sowohl auf Mikrofilm wie mittels PDF/a „archivierbar“ wären, was dann Datenträger wie M-Disc und GlassDisc einschließt. Problematisch und zeitaufwendig erscheint allerdings nur die Redigitalisierung beim Mikrofilm. Nach dem 3-2-1-Backup spricht aber nichts dagegen, Archive doppelt mit langsamen und schnellen Speichern zu pflegen. Dann könnte man zudem auch Festplatten neben den LZ-tauglichen Datenträgern M-Disc und GlassDisc für kurzfristige (Zwischen-)Archivierung heranziehen (Ullrich [1], 2009, S.10.15) .

Um die einleitende Aussage Stolls zu beantworten, können die Feststellungen dahingehen interpretiert werden, dass dieser nur teilweise recht hatte. Digitale Daten sind zwar nicht LZ-archivierbar, allerdings nur mit den derzeitigen Mitteln nicht. Stehen entsprechende Datenträger zur Verfügung und erweist sich die 5D Glass Disc als tauglich, so verhält es sich mit digitalen Daten nicht anders als mit physischen Objekten anderer Art. Wenn man auf sie Acht gibt, überdauern sie, wenn nicht, kann man sie natürlich auch zerstören. Je mehr verteilte Lagerorte zudem zum Einsatz kommen (bei Hadoop das Äquivalent: Nodes), desto geringer wird auch die Wahrscheinlichkeit eines Totalverlustes. Konkret bedeutet hier nämlich der Untergang eines einzelnen Archives durch Naturkatastrophen, kriegerische Konflikte oder auch durch technisches oder menschliches Versagen nicht den Untergang des Backup-Archives als Ganzes. Aus jüngster Zeit zeugt der Einsturz des Kölner Stadtarchives jedoch davon, dass man bei fehlerhafter Strategie dennoch Daten verlieren kann (WDR, 2009). Facebook und Google machen es elektronisch vor, für KMUs und Private bietet sich die Einlagerung in Bergwerkstollen wie dem „earthDATAsafe“ an, um Datenträger zuverlässig aufzubewahren. Regierungsorganisationen betreiben hier auch eigene Bunker, so z.B. die Österreichische Bundesregierung in Salzburg, wo das Bundesrechenzentrum sein Backuparchiv hat.

Ungelöst muss allerdings die Frage nach der richtigen LZ-Archivierungsstrategie bleiben. Hier gehen die Meinungen auseinander. Einige setzen auf Emulation, andere auf Migration. Am zuverlässigsten wäre zwar OAIS-konform die Migration anzusehen, da hier aktuelle Techniken zum Einsatz kommen. Wenn diese jedoch nicht die Nutzung des Contents gewährleisten kann, ist die nächstbeste Strategie die Emulation. Wie allerdings gezeigt wurde, haben letztlich beide Strategien ihre Grenzen. Allerdings kann man dies auch so interpretieren, wie beim doppelten Vorrätighalten von unterschiedlichen Datenträgern beim 3-2-1-Backup, einfach beide Strategien

anzuwenden. Letztlich bleibt aber zu sagen, dass es am zuverlässigsten ist, Bauanleitungen für Hardware zu archivieren. Dafür stehen mit PDF/a und Mikrofilm, sowie langlebigen Datenträgern durchaus auch heute schon die Mittel zur Verfügung. Bitstreams lassen sich durch obige Strategien ebenfalls über die Zeit retten. Die Hardware jedoch muss, wenn sie gebraucht wird, anhand dieser archivierten Bauanleitungen nachkonstruiert werden, wenn Emulation und Migration versagen. Was nämlich durch Menschenhand einmal gebaut wurde, kann anhand von Anleitungen à la Kochrezepten jederzeit wieder erfolgen. Nur müssen diese „Kochrezepte“ eben verfügbar sein. Genannt wurde hier nicht einmal so abwegige Science Fiction des Raumschiffes Enterprise mit dem Replikator. Ohne Bauanleitung für die Zusammensetzungen kann auch dieser nichts replizieren. Daher heute archivieren!

10 Literatur

- Bárány, B. (2006). Informationsverlust durch Digitalisierung. Grundlagen und Konzepte zur Langzeitsicherung digitaler Informationen. Saarbrücken: VDM Verlag.
- Black-Veldtrup, M., Meusch, M., Przigoda S. (2002). Zugänglichkeit verbessern: Das DFG-Projekt „Entwicklung von Werkzeugen zur Retrokonversion archivischer Findmittel“. *Der Archivar*, 2002 (2), 111-117.
- Betz, L., Widhalm, T. (2016). *Icinga 2*. dpunkt, Hedeilberg.
- Borghoff, U. M., Rödiger, P., Scheffczyk, J., Schmitz, L. (2005). *Long-Term Preservation of Digital Documents*. Heidelberg: Springer.
- Brandt, O. (2009): PREMIS. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (6.9-6.13)*. Boizenburg: Verlag Werner Hülsbusch.
- Brodersen, M., Schrimpf, S. (2009). Langzeitarchivierung von elektronischen Publikationen durch die Deutsche Nationalbibliothek. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (18.3-18.12)*. Boizenburg: Verlag Werner Hülsbusch.
- Brübach, N. (2009). Das Referenzmodell OAIS. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (4.3-4.14)*. Boizenburg: Verlag Werner Hülsbusch.
- Burleson Consulting (o.D.). SQL Server vs. Oracle Datatypes. Abgerufen von URL: http://www.dba-oracle.com/t_migrating_sql_server_vs_oracle_datatypes.htm
- Dobratz, S., Schoger, A. (2009). Vertrauenswürdigkeit von digitalen Langzeitarchiven. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (5.1-5.7)*. Boizenburg: Verlag Werner Hülsbusch.
- Egger, A. (2005). *Trusted Digital Repositories - Architektur und Design*. Dissertation am Institut Integrierte Studieren, Johannes Kepler Universität, Linz.
- Enders, M. (2009). Bilddokumente. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (17.8-17.18)*. Boizenburg: Verlag Werner Hülsbusch.

- Funk, S. E. (2009). Emulation. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (8.16-8.23). Boizenburg: Verlag Werner Hülsbusch.
- Gravenhorst, C. (2007). METAe - The „Metadata Engine“ Project - New technologies for automated conversion of historic treasures. In: o.V. (2007). Digitalisieren - international Projekte in Bibliotheken und Archiven. Berlin: BibSpider.
- Hofmann, R., Wiesner, H. J. (2015). Bestandserhaltung in Archiven und Bibliotheken. Berlin, Wien, Zürich: Beuth Verlag GmbH.
- Hughes, L. M. (2004). Digitizing collections. Strategic issues for the information manager. London: Facet Publishing.
- Huth, K. (2009). Computermuseum. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (8.24-8.31). Boizenburg: Verlag Werner Hülsbusch.
- Huth, K. (2009). Computerspiele. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (17.116-17.130). Boizenburg: Verlag Werner Hülsbusch.
- Huth, K. (2009). Textdokumente. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (17.3-17.7). Boizenburg: Verlag Werner Hülsbusch.
- Jehn, M., Schrimpf, S. (2009). Bibliotheken. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (2.6-2.8). Boizenburg: Verlag Werner Hülsbusch.
- Kapsch.Net (2017). Das sicherste Datenzentrum Österreichs. Abgerufen von URL: <https://www.kapsch.net/kbc/solutions/Operating/high-securitycomputingcenter?lang=de-AT>
- Keitel, C. (2009). Mikroverfilmung. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (8.32-8.33). Boizenburg: Verlag Werner Hülsbusch.
- Keitel, C., Schoger A. (Hrsg.) (2013). Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644. Berlin, Wien, Zürich: Beuth Verlag GmbH.
- Larson, R, Carbone, J. (2009). Microsoft Windows Server 2008 Hyper-V. Unterschleißheim: Microsoft Press.

- Lazinger, S. S. (2001). *Digital Preservation and Metadata. History, Theory, Practice*. Englewood: Libraries Unlimited, Greenwood Publishing Group.
- Landesarchiv Nordrhein-Westfalen (2015): Leitfaden für langzeitstabile Datenformate in der elektronischen Aktenführung. Abgerufen von URL: http://www.archive.nrw.de/lav/abteilungen/fachbereich_grundsaeetze/BilderKartenLogosDateien/Behoerdeninformation/Bausteine_III_Langzeitstabile_Datenformate.pdf
- Library of Congress (2000). *A digital strategy for the Library of Congress*. Washington: National Academy Press.
- Ludwig, J. (2009). Auswahlkriterien. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (7.9-7.11)*. Boizenburg: Verlag Werner Hülsbusch.
- Luther, J. (2004). Backup mit Konzept. Abgerufen von URL: <http://www.tecchannel.de/a/backup-mit-konzept,402355,8>
- Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.) (2009). *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung Version 2.0*. Boizenburg: Verlag Werner Hülsbusch.
- Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.) (o.D.). *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung Version 2.3*. Abgerufen von URL: http://eprints.rclis.org/28088/1/nestor-handbuch_23.pdf
- Neuroth, H., Strathmann, S., Oßwald, A., Ludwig, J. (Eds.) (2013). *Digital Curation of Research Data. Experiences of a Baseline Study in Germany*. Glückstadt: Verlag Werner Hülsbusch.
- ORF [0] (2016). E-Medikation: Verordnung erst, wenn Software funktioniert. Abgerufen von URL: <http://orf.at/stories/2361274>
- ORF [1] (2016). Web befreit sich von Altlasten. Abgerufen von URL: <http://orf.at/stories/2354474>
- ORF [2] (2016). Kern vermisst Kaktus in Regierungsbunker. Abgerufen von URL: <http://oesterreich.orf.at/stories/2820282>
- ORF [3] (2016). Weltweit sicherste Cloud aus Linz. Abgerufen von URL: <http://ooe.orf.at/news/stories/2810865>
- Oßwald, a. (2009). Das Referenzmodell OAIS – Open Archival Information System. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (4.1-4.2)*. Boizenburg: Verlag Werner Hülsbusch.

- PC-Welt (2001): Kopierschutz. So funktioniert er. Abgerufen von URL: <http://www.pcwelt.de/ratgeber/Kopierschutz-So-funktioniert-er-67529.html>
- PDF Tools AG. (2009). White Paper. PDF/A - der Standard für die Langzeitarchivierung. Version 2.4. Abgerufen von URL: <https://www.pdf-tools.com/public/downloads/whitepapers/whitepaper-pdf-a.pdf>
- Potenzialfinder (2016). Wissenspyramide. Abgerufen von URL: <http://www.potenzialfinder.com/de/leistungen/wissensbilanz/glossar.html?glossarid=14>
- Rathje, U. (2002). Technisches Konzept für die Datenarchivierung im Bundesarchiv. *Der Archivar*, 2002 (2), 117-120.
- Rieder, B. (2013). Ratgeber Langzeitarchivierung: Dateiformate und Speichermedien. Lebensdauer von Archivierungsmedien. Abgerufen von URL: <http://www.tecchannel.de/a/ratgeber-langzeitarchivierung-dateiformate-und-speichermedien,2039663,7>
- Rothenberg, J. [0] (1999). Ensuring the Longevity of Digital Information. Abgerufen von URL: <https://www.clir.org/pubs/archives/ensuring.pdf>
- Rothenberg, J. [1] (1999). Using Emulation to Preserve Digital Documents. Abgerufen von URL: <http://www.imaginar.org/taller/dppd/DPPD/148%20pp%20usingemulation.pdf>
- Spindler, G., Hillegeist, T. (2009). Langzeitarchivierung wissenschaftlicher Primärdaten. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (16.14-16.23)*. Boizenburg: Verlag Werner Hülsbusch.
- Stoll, C. (1998). *Die Wüste Internet. Geisterfahrten auf der Datenautobahn*. Frankfurt am Main: Fischer Taschenbuch Verlag GmbH.
- Syylex (2011). GlassMasterDisc. Abgerufen von URL: http://www.syylex.com/tl_files/syylex/Downloads/Flyer_SYYLEX_EN_G.pdf
- Tischer, M., Jennrich, B. (1998): *CD-Brenner*. Düsseldorf: Data Becker.
- Ullrich, D. (2009). Bitstream Preservation. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (10.15-10.18)*. Boizenburg: Verlag Werner Hülsbusch.
- Ullrich [1], D. (2009). Digitale Speichermedien. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), *nestor Handbuch. Eine*

- kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (10.6-10.18). Boizenburg: Verlag Werner Hülsbusch.
- Vlaeminck, S. (2009). Organisation. In: Neuroth, H., Oßwald, A., Scheffel, R., Strathemann, S., Jehn, M. (Hrsg.), nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0 (10.15-10.18). Boizenburg: Verlag Werner Hülsbusch.
- Vincent, J. (2016): 'Five-dimensional' glass discs can store data for up to 13.8 billion years. Abgerufen von URL:
<http://www.theverge.com/2016/2/16/11018018/5d-data-storage-glass>
- von Suchodoletz, D. (2009): Funktionale Langzeitarchivierung digitaler Objekte. Erfolgsbedingungen des Einsatzes von Emulationsstrategien. Göttingen: Cuvillier.
- WDR (2009). Einsturz des Kölner Stadtarchivs. Abgerufen von URL:
<http://www1.wdr.de/archiv/stadtarchiv-ubahn/index.html>
- White, T. (2012). Hadoop. The Definitive Guide. Sebastopol: O'Reilly
- Windeck, C. (2013). Bitfäule. Tücken bei Speicherung und Übertragung großer Datenmassen. c't, 2013 (21), 176-179.

Abbildungsverzeichnis

Abbildung 1: Was Bits aus Nullen und Einsen darstellen können.....	4
Abbildung 2: Digitale Bitfäule und ihr Ergebnis.....	5
Abbildung 3: Die Wissenspyramide	8
Abbildung 4: schnelle und Langsame / geringe und hoch kapazitative Speicher	14
Abbildung 5: Das 3-2-1-Backup	16
Abbildung 6: Das Funktionsmodell des OAIS	37

Tabellenverzeichnis

Tabelle 1: Haltbarkeit von Datenträgern	7
Tabelle 2: Restore mit wenigen Mediensätze.....	17
Tabelle 3: wichtige Dateisysteme.....	56

Abkürzungen

AV	Audio / Video
d.h.	das heißt
DMS	Dokumentenmanagement-System
Eds.	Editors
etc...	et cetera
EB	Exabyte
f.	folgende
ff.	fortfolgende
GB	Gigabyte
Hrsg.	Herausgeber
idR	in der Regel
khz	Kilohertz
LZ / lz	langzeit
LZA	Langzeitarchivierung
max.	maximal
Mrd.	Milliarden
NTFS	New Technology File System
OAIS	Open Archival Information System
OS	Operating System (Betriebssystem)
o.V.	ohne Verfasser
RAID	Redundant Array of Independent Discs
sog.	so genannte(r)
TB	Terabyte
u.a.	unter anderem
u.u.	unter Umständen
usw...	und so weiter
v.	von
v.a. / V.a.	vor allem
z.B.	zum Beispiel

Anhang

Wichtige Dateisysteme

Name	Hersteller	System	max. Dateigröße	max. Partitionsgröße	max. Dateinamenlänge	Metadaten	Journaling	Anmerkungen
exFAT	Microsoft	Windows (theoretisch Linux, Mac-OS)	16 EB	512 TB	255 Zeichen	nein	nein	ungeeignet als Austauschformat
Ext2	Remy Card u. a.	Linux, BSD, (Windows und Mac mit Zusatztreiber)	16 GB bis 2 TB	2 TB bis 32 TB	255 Zeichen	ja	nein	veralteter Linux-Standard
Ext3	Stephen Tweedie u. a.	Linux, BSD, (Windows und Mac mit Zusatztreiber)	16 GB bis 2 TB	2 TB bis 32 TB	255 Zeichen	ja	ja	lange Linux-Standard
Ext4	Kollektiv	Linux, (Mac-OS mit Zusatztreiber)	16 GB bis 16 TB	2TB bis 32 TB	255 Zeichen	ja	ja	aktueller Linux-Standard
BTRFS	Oracle	künftige Linuxdistributionen	16 EB	16 EB	255 Zeichen	ja	ja	evtl. künftiger Linux-Standard
JFS	IBM	Linux, OS/2	4 PB	32 PB	255 Zeichen	ja	nur Metadaten	Linux-Live-Systeme oft ohne JFS-Unterstützung
Reiser FS	Namesys (Firma des Namensgebers)	Linux, BDS, (Windows mit Zusatztreiber)	4-8 GB	16 TB	255 Zeichen	ja	ja	wird nicht weiterentwickelt
ReFS	Microsoft	Windows 8 Server, künftige Windows-Desktop-Versionen	16 EB	16 EB	255 Zeichen	ja	ja	künftiger Windows-Standard
FAT32	Microsoft	Windows, Linux, OS X	4 GB	2 TB	256 Zeichen	nein	nein	einfaches Austauschformat
NTFS	Microsoft	Windows, Linux, OS X	16 EB	16 EB	255 Zeichen	ja	ja	Windows Standard

Tabelle 3: wichtige Dateisysteme

Quellen der Bilder, die nicht in „yed“ für Das 3-2-1-Backup-Schema verfügbar waren. (Abbildung 4) / <https://www.yworks.com/products/yed>:

- Tape Roboter:
<http://melauspartners.com/backup-tape-library>
- NAS:
<http://www.bhphotovideo.com/explora/computers/tips-and-solutions/nas-servers-explained>
- Safe:
<http://www.iconarchive.com/show/security-icons-by-aha-soft/safe-icon.html>
- Bank-Schließfach:
<https://www.raikaerberndorf.at/Privatkunden/SparenAnlegen/Services/Sparbuch-Schliessfach/c923.html>

Eidesstattliche Erklärung

Hiermit erkläre ich ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig angefertigt, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Ich erkläre außerdem, dass die vorliegende Arbeit bei keiner anderen Institution (Fachhochschule, Universität, Pädagogische Hochschule oder vergleichbare Bildungseinrichtung) zur Erlangung eines akademischen Grades eingereicht wurde.

Ort, Datum

Unterschrift