nestor

Vertrauenswürdige und abgesicherte Langzeitarchivierung multimedialer Inhalte

Andrea Oermann, Gerald Jäschke, Jana Dittmann
Otto-von-Guericke-Universität Magdeburg

nestor-materialien 14

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Vertrauenswürdige und abgesicherte Langzeitarchivierung multimedialer Inhalte

Andrea Oermann
Gerald Jäschke
Jana Dittmann

Otto-von-Guericke-Universität Magdeburg

nestor-materialien 14

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Herausgegeben von

nestor - Kompetenznetzwerk Langzeitarchivierung und
Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland

nestor - Network of Expertise in Long-Term Storage of Digital Resources

http://www.langzeitarchivierung.de

Projektpartner:

Bayerische Staatsbibliothek, München
Bundesarchiv
Deutsche Nationalbibliothek (Projektleitung)
FernUniversität in Hagen
Humboldt-Universität zu Berlin - Computer- und Medienservice / Universitätsbibliothek
Institut für Museumsforschung, Berlin
Niedersächsische Staats- und Universitätsbibliothek, Göttingen

Betreuer dieser Veröffentlichung:

Humboldt-Universität zu Berlin
Susanne Dobratz

Projektkoordination:

Otto-von Guericke-Universität Magdeburg
Prof. Dr. Jana Dittmann

URN: urn:nbn:de:0008-2009081423
http://nbn-resolving.de/urn:nbn:de:0008-2009081423

Vorwort

Die vorliegende Expertise analysiert den Stand derzeit existierender digitaler Archive, die sich auf die langfristige Bestandserhaltung und Bereitstellung multimedialer Daten, vor allem jedoch auf Ton- und Videodaten spezialisiert haben. Damit stellen die Autoren einen wichtigen Baustein bereit, um die existierenden Kriterienkataloge und die darin formulierten Anforderungen auf das Gebiet der Langzeitarchivierung multimedialer Objekte zu übertragen, das im Rahmen der zweiten Projektphase des nestor-Kompetenznetzwerkes in den Vordergrund gerückt wurde.

Die im nestor Kriterienkatalog aufgestellten abstrakten Anforderungen an vertrauenswürdige digitale Langzeitarchive werden unter dem Gesichtspunkt der Bewahrung und Benutzbarerhaltung kontinuierlicher digitaler Medien diskutiert und verfeinert. So lassen sich aus den abstrakten Anforderungen messbare Parameter und konkrete Maßnahmen für den Umgang mit diesen speziellen Medien innerhalb ihres jeweiligen Kontextes ableiten. Dies unterstützt die praktische Anwendbarkeit des bestehenden nestor-Kriterienkatalogs.

Den zweiten Schwerpunkt der Untersuchungen bildet die detaillierte Analyse von sicherheitsrelevanten Eigenschaften und Anforderungen. Damit wird in einer einzigartigen Weise ausführlich der Bezug der Eigenschaften der IT-Sicherheit zum Thema Vertrauenswürdigkeit und digitale Langzeitarchive hergestellt. Die durch die Autoren dieser Expertise herausgearbeiteten Detailanforderungen an die Sicherheitseigenschaften digitaler Archive und digitaler Objekte und die Maßnahmenkataloge sind von hoher Praxisrelevanz für den Aufbau zukünftiger Medienarchive.

Als besonders wichtig für die weitere Arbeit im Rahmen des Kompetenznetzwerkes nestor erachte ich den im Fazit dieser Studie dargelegten und durch die vorangegangenen Detailanalysen belegten Standardisierungsbedarf in allen Bereichen, die den Aufbau vertrauenswürdiger und sicherer digitaler Langzeitarchive betreffen. Das bezieht sich sowohl auf organisatorische als auch auf technische Aspekte. Hieraus lässt sich eine Roadmap für die weitere Arbeit im Rahmen des DIN-Normenausschusses „Bibliotheks- und Dokumentationswesen“ (NABD 15) ableiten.

Die Erstellung der Expertise war geprägt durch eine sehr intensive und anregende Zusammenarbeit, so dass bereits während des Arbeitsprozesses wichtige Aspekte in die aktuelle Arbeit der nestor-Arbeitsgruppe „Vertrauenswürdige digitale Archive“ Eingang finden konnten.

Für die Partner des Projektes nestor – Kompetenznetzwerk Langzeitarchivierung

Susanne Dobratz

Humboldt-Universität zu Berlin

Aufgabenstellung der Expertise

Innerhalb der Studie werden Ansatzpunkte für die Nutzung von Langzeitarchivierungstechnologien für multimediale Kollektionen, wie sie z.B. in Musikarchiven, Rundfunkarchiven, Medienzentren von Hochschulen, multimedialen e-Learning-Kollektionen usw. existieren, untersucht.

Dies wird aufbauend auf den Ergebnissen der nestor-Expertise „Perspektiven der Langzeitarchivierung multimedialer Objekte" und der Studie „Integration von Archiv-Metadaten in einem zentralisierten Metadaten-Repository" des Institutes für Rundfunktechnik mit dem Ziel erarbeitet, zunächst einen Überblick über diesen Anwendungsbereich zu geben.

Dabei ist es von besonderer Bedeutung für die diversen Medienproduzenten, die Menge an anfallenden Daten und deren technische Aufbereitung sowie die zugrunde liegenden Kriterien zu erheben. Zu diesem Zweck sind gezielte Umfragen in Abstimmung mit dem Auftraggeber vorzubereiten und durchzuführen.

Die zu erstellende Studie untersucht weiterhin die Anforderungen und die grundsätzliche technische Eignung einer Anwendung von bereits am Markt existierenden Systemen zur Langzeitarchivierung von Multimediakollektionen (aufbauend auf der Nestor-Expertise „Vergleich bestehender Archivsysteme") sowie hinsichtlich der bereits bei den diversen Archiven im Einsatz befindlichen Systeme für die Archivierung.

Zu diesem Zweck werden innerhalb der Erhebungsaktivitäten auf der Grundlage von Interviews systematisch die technischen Infrastrukturen und die Eigenschaften der Plattformen exemplarischer Archive von ARD-Mitgliedsanstalten sowie Medienzentren an Hochschulen untersucht. Dabei sollen einerseits die verwendeten technischen Plattformen und Mechanismen zur Speicherung von hochkomplexen und verteilt liegenden Inhalten, Inhaltsbeschreibungen und Rechte-Daten sowie Metadaten zur Langzeitarchivierung erfasst werden. Einen wichtigen Aspekt der Studie bildet die Erfassung von Einsatzmöglichkeiten von Sicherheitstechnologien, wie digitale Signaturen und Wasserzeichen im Multimedia-langzeitarchivierungsumfeld.

Andererseits sollen die organisatorischen Rahmenbedingungen, wie im „nestor Kriterienkatalog Vertrauenswürdige Digitale Archive" beschrieben, erfasst werden. Damit sollen die bisher von nestor in dem Bereich „Vertrauenswürdige Langzeitarchive" erarbeiteten Grundlagen an praktischen Beispielen validiert werden.

Zugleich soll daraus abgeleitet werden, in welchen Bereichen es deutlichen Standardisierungsbedarf gibt, in denen sich das Kompetenznetzwerk Langzeitarchivierung als Ganzes engagieren und Aktivitäten initiieren muss.

In der Studie sollen über existierende Werkzeuge, Architekturen und Lösungen hinaus auch hochgradig verteilte Speicherlösungen (OAIS als Ausgangs-Architektur) betrachtet werden, um z.B. Potenziale für den Einsatz von Grid- und anderen Virtualisierungstechnologien auf allen Ebenen aufzuzeigen.

Inhaltsverzeichnis

1 Einleitung, Motivation und Überblick

Die *Langzeitarchivierung* dient grundsätzlich der Erhaltung von Information von bleibendem wissenschaftlichem, künstlerischem oder gesellschaftlichem Wert [BRSS03]. Dabei sollen der Zugriff und die Verfügbarkeit für eine bestimmte autorisierte Zielgruppe sowohl in der Gegenwart als auch in der Zukunft über einen Zeitraum hinaus gewährleistet sein, innerhalb dessen technologische und soziokulturelle Veränderungen eintreten werden. Langzeit bedeutet die verantwortliche Entwicklung von Strategien, die den beständigen, vom Informationsmarkt verursachten Wandel bewältigen können [ScLi04]. Laut *OAIS-Referenzmodell* [CCSDS02], auf welches sich diese Studie grundlegend bezieht, bezeichnet ein Archiv eine Organisation zusammengesetzt aus Personen und Systemen, deren Aufgabe bzw. Verantwortung es ist, die Information zu erhalten und sie für eine bestimmte Zielgruppe zugänglich zu machen. Für ein Langzeitarchivierungssystem gilt darüber hinaus die Anforderung, dass Information über einen längeren Zeitraum sicher aufbewahrt und zugänglich gemacht werden muss, wobei die Zuordenbarkeit der archivierten Daten zu jedem Zeitpunkt sichergestellt werden muss [BRSS03]. Eine Information soll also auch in 100 Jahren noch verfügbar und lesbar sein.

> *„Langzeit" ist die Umschreibung eines nicht näher fixierten Zeitraumes, währenddessen wesentliche nicht vorhersehbare technologische und soziokulturelle Veränderungen eintreten, die sowohl die Gestalt als auch die Nutzungssituation digitaler Ressourcen in rasanten Entwicklungszyklen vollständig umwälzen werden. Dabei spielen nach bisheriger Erfahrung das Nutzerinteresse der Auf- und Abwärtskompatibilität alter und neuer Systemumgebungen eine Rolle nur dann, wenn dies dem Anbieter für die Positionierung am Markt erforderlich scheint. „Langzeit" bedeutet für die Bestandserhaltung digitaler Ressourcen nicht die Abgabe einer Garantieerklärung über fünf oder fünfzig Jahre, sondern die verantwortliche Entwicklung von Strategien, die den beständigen, vom Informationsmarkt verursachten Wandel bewältigen können. [ScLi04]*

Während die Langzeitarchivierung generell das Erhalten jeglicher Information egal welchen Formats, einschließlich der dazugehörigen Methoden bezeichnet, bezieht sich die *digitale Langzeitarchivierung* auf ausschließlich Methoden und Strategien zur Erhaltung digitaler Information. Digitale Information unterliegt einem vom Informationsmarkt verursachten sehr schnellen Wandel. Ein heute gängiges Format ist morgen bereits überholt. Aufgrund der Schnelllebigkeit der technischen Weiterentwicklungen steht die Langzeitarchivierung vor neuen Herausforderungen und Anforderungen. Um Zugang zur Information zu erlangen wird immer ein Abspielsystem [BRSS03] benötigt, das die in Zeichenströmen gespeicherte Information interpretiert und darstellt. Die digitale Langzeitarchivierung muss sicherstellen, dass die Formate immer interpretierbar sind und der Zugang über ein benötigtes Abspielsystem gewährleistet ist. Dazu werden verschiedene Strategien verfolgt wie das z.B. Gewährleisten der Auf- und Abwärtskompatibilität alter und neuer Systemumgebungen oder das Migrieren veralteter Formate, in denen die Information gespeichert ist, in neuere, in einer Systemumgebung abspielbare bzw. interpretierbare Formate.

In Bezug auf digitale Information sind heutzutage zwei Phänomene vorhanden [ScLi04]:

- **Digitalisate:** Zunehmender Umfang von Digitalisierungen (Konvertierungen) von ursprünglich in analoger Form vorliegender Information einschließlich ihrer Erschließung und Bereitstellung.

- **Born Digital:** Ständig anwachsende Menge und Heterogenität von originär in digitaler Form vorliegender Information.

Motivationen für die Digitalisierung bestehenden Materials sind die Rettung vor Verfall als auch die Vereinfachung und orts- sowie zeitunabhängige Verteilung der Verteilung der Benutzerzugriffe über Datennetze. Die wachsende Menge ausschließlich digital vorliegender Information zieht eine wachsende Relevanz dieser Information als Bestandteil unserer kulturellen Überlieferung sowie die Bedeutung ihrer dauerhaften Verfügbarkeit für Wissenschaft und Forschung nach sich. Deren Langzeiterhaltung und Langzeitverfügbarkeit ist damit unumgehbar, nicht nur um einen Verlust der Ressourcen zu verhindern, sondern auch um konkurrenzfähig zu bleiben.

Medien und Multimedia

Die zu archivierende digitale Information kann in den *verschiedensten Medien* dargestellt sein, wie etwa Text (Dokumente), Bild (Grafiken), Audio (Tonaufnahmen) oder Video (Bild und Ton). In heutigen digitalen Archiven liegt vorrangig *Multimedia*-Information vor, d.h. Information zusammengesetzt aus unterschiedlichen Medien, wie z.B. Video. Originär analog vorliegende Textdokumente werden eingescannt und als Bild zusammen mit den in der Erschließung erstellten Metadaten im Archiv gespeichert. Weiterhin werden riesige Mengen an Multimediainformationen digital produziert wie z.B. in Hochschul-Medienzentren oder in Rundfunk- und Fernsehanstalten. Für die vorliegende Expertise sollen diese beiden Szenarien im späteren Verlauf als Schwerpunkte und Beispiele dienen.

Die Bedeutung des Wortes Medien variiert je nach Kontext in welchem der Begriff gebraucht wird. Im Zusammenhang der digitalen Langzeitarchivierung multimedialer Inhalte soll die folgende Definition gelten: *Medien* verteilen und repräsentieren Information [StNa02]. So gibt es unterschiedliche Medien wie z.B. Text, Grafiken, Bilder, Stimme und Sprache, Musik und Töne. Medien haben demnach eine Entsprechung in den menschlichen Sinnen, wie wir Information wahrnehmen (sehen, hören, tasten, usw.).

Kriterien zur Unterscheidung von Aspekten in Bezug auf Medien sind wie folgt festgelegt [MHEG93]:

- **Wahrnehmungsmedien** Wie nehmen Menschen Information wahr?
- **Repräsentationsmedien** Wie ist die Information im Computer encodiert?
- **Präsentationsmedien** Welches Medium realisiert den Output/ Input der Information aus/ in den Computer?
- **Speicherungsmedien** Wo ist Information gespeichert?
- **Transmissionsmedien (Kabel, Satellit, Funk-Radiowellen)** Welches Medium verbreitet/ überträgt die Information?
- **Informationsaustauschmedien** Mit welchem Datenmedium wird die Information ausgetauscht zwischen zwei unterschiedlichen Orten?

Entsprechend des Darstellungsraumes und der Dimensionen können Medien in diskrete (zeitunabhängige) wie Text und Bilder und kontinuierliche (zeitabhängige) Medien wie Audio und Video unterschieden werden. Weiterführend ist ein *Multimediasystem* eine Komposition aus mehr als einem Medium, wobei der Computer als Repräsentationsmedium eingeschlossen ist [StNa02].

Heutzutage gibt es unzählige unterschiedliche Formate und Medien, vielmehr noch Multimedia, unterschiedliche Versionen von Präsentationssoftware und Systemsoftware und variierende Hardware. Aufgrund verschiedener Standards und Entwicklungsgruppen ist eine Unverträglichkeit der Formate oftmals alltäglich. Sowohl die Formate als auch die Software unterliegen einem ständigen Wandel. Ein zunehmender Umfang digitaler Datenträger mit unterschiedlichen Alterungszyklen ist zu verzeichnen. So steht man immer wieder vor der Entscheidung welches Medium für die Langzeitarchivierung digitaler Information am sinnvollsten ist.

Wie eingangs erwähnt, besteht die Problematik im Kontext digitaler Langzeitarchivierungssysteme nun darin, dass Daten als Bitströme (digitale Zeichenströme) gespeichert sind und ein passendes *Abspielsystem* [BRSS03] benötigen, um zugänglich zu sein. Ein Abspielsystem dient der Zusammensetzung der Zeichenfolgen und Bausteine eines definierten Alphabets in einer geeigneten Umgebung. Ein Abspielsystem ist aus verschiedenen miteinander interagierenden Komponenten zusammengesetzt. Dazu zählen Komponenten der Hardware, der Systemsoftware (Betriebssystem, Treibersoftware, usw.) und der Präsentationssoftware (Editor, Browser, usw.). Die Anwendung der geeigneten Präsentationssoftware ist abhängig von der Präsentationsform der anzuzeigenden Information, dem Format.

Umdenken – Langzeitarchivierung als Prozess

In Bezug auf die digitale Langzeitarchivierung muss demnach umgedacht werden. Die Langzeitarchivierung als Bestandserhaltung muss als ein dynamischer Prozess der Speicherung gehandhabt werden. Eine Sicherung des Fortbestandes ist nur durch aufwändige Erhaltungsmaßnahmen möglich. Um einen Informationsverlust zu verhindern, müssen geeignete Maßnahmen für die Erstellung von Sicherheitskopien, vor allem aber auch für die Umkodierung zur Anpassung an neue Geräte und Systemsoftware entwickelt werden. Darüber hinaus sind neben der eigentlichen Informationsspeicherung (Objektspeicherung) auch die Erstellung und Pflege von Metadaten notwendig. Geeignete und umsetzbare Backup- bzw. Erhaltungsstrategien wie fortgesetztes Umkopieren, Migration, reversibler Übergang zu neuen Codes oder Emulatorprogramme müssen entwickelt werden, um digitale Zeichenströme sicher zu konservieren.

In digitalen Systemen, wie sie in der digitalen Langzeitarchivierung zum Einsatz kommen, steht man neuen Bedrohungen gegenüber, die nicht nur den schleichenden, oft unbemerkten physischen Zerfall des Mediums betreffen. So ist die schnelle Überalterung der interpretierenden Technik ebenso ein Problem, wie Gewährleistung der Verfügbarkeit authentischer und integerer Information, die Sicherstellung der Vertraulichkeit und die Zugriffsregelung.

Vertrauenswürdige und abgesicherte digitale Langzeitarchivierungssysteme

Ein vertrauenswürdiges digitales Langzeitarchivierungssystem muss technisch, organisatorisch und rechtlich abgesichert sein. Diese drei Aspekte sind wie folgt zu unterteilen: [Neu05]

Organisatorische Aspekte

Wer übernimmt wann, wofür und wie lange die Verantwortung? Dazu zählen:

- Sammel- und Auswahlkriterien
- Erhaltungsrichtlinien (*Preservation Policies*)
- Kooperationen
- Standardisierungen

Technische Aspekte

Welches sind Bestandserhaltungsmaßnahmen bzw. Strategien zur langfristigen Verfügbarkeit digitaler Objekte? Dazu zählen:

- **Strategie**
  - *Migration* Transformation eines digitalen Objektes von einer technischen Umgebung in eine andere.
  - *Emulation* Nachbildung der für die Benutzung digitaler Objekte erforderlichen originalen technischen Umgebung.
- **Methodik**
  - *OAIS-Referenzmodell*

- **Metadaten** Metadaten sollen sicherstellen, dass Information langfristig interpretiert und genutzt werden kann. Unterteilung der Metadaten in:
  - *Technische Metadaten*: Beschreiben die erforderliche technische Umgebung *Metadaten zu einzelnen Dateitypen*
  - *Strukturelle Metadaten*: Beschreiben Relationen zu anderen Objekten im Archiv.
  - *Administrative Metadaten*: Beschreiben Lebenszyklus des Objekts und bestimmen Archivierungsmaßnahmen.
  - *Rechtliche Metadaten*: Beschreiben Zugriffsmodalitäten auf Objekte.
  - *Beschreibende Metadaten*
- **Persistent Identifier (PI)** Persistente Identifikatoren identifizieren dauerhaft eine Ressource eindeutig und dienen deren langfristigen Auffindbarkeit und damit zu deren langfristigen Benutzbarkeit.
- **Technische Infrastruktur** Eine geeignete, den Anforderungen angepasste technische Infrastruktur sollen die langfristige Verfügbarkeit garantieren. Sie beinhaltet:
  - Medien, Formate
  - Systemumgebungen, Hard- und Software
  - Aufbau und Vernetzung der Hardware
  - Zugriffsregelungen
  - File Format Registries
  - Werkzeuge zur automatisierten Formatidentifikation, Formatvalidierung und Formatcharakterisierung.
- **Sicherheitsaspekte** Angemessene Realisierung der Integration von Sicherheitsmechanismen zur Sicherung der Integrität, Authentizität und Verfügbarkeit der zu archivierenden Information.

Rechtliche Aspekte

Welche Rechtsgrundlagen (national/ international) müssen bei der langfristigen Archivierung von Information beachtet werden? Dazu zählen:

- Urheberrecht
- Verwertungsrecht
- Pflichtexemplarrecht
- Archivrecht
- Datenschutz

und weitere.

Während die rechtlichen Aspekte in der Expertise nestor - Digitale Langzeitarchivierung und Recht [Nes04] erörtert wurden, sind Kriterien zur Bewertung der Vertrauenswürdigkeit eines digitalen Langzeitarchivs sowohl in organisatorischer als auch in technischer Hinsicht im nestor-Kriterienkatalog [Nes06] vertrauenswürdige digitale Langezeitarchive definiert. Die Festlegung der Kriterien geschah in engem Kontakt sowohl mit unterschiedlichsten Gedächtnisorganisationen und Produzenten von Information als auch weiteren Betroffenen und Experten. Mit diesem offenen Vorgehen wurden eine hohe Allgemeingültigkeit und Praxistauglichkeit sowie eine breite Akzeptanz der Ergebnisse beabsichtigt.

Im nestor-Memorandum [Nes06a] sind Empfehlungen festgehalten bzgl. pro-aktiver Maßnahmen zur Erhaltung der Langzeitverfügbarkeit digitaler Information, ohne die Gefahr drohen würde, dass wichtige Kulturgüter verloren gehen. Diese Empfehlungen beschreiben die Rahmenbedingungen für eine nationale Langzeitarchivierungs-Policy und sind unterteilt in die folgenden Bereiche:

- Verantwortung für die Langzeiterhaltung digitaler Information
- Auswahl, Verfügbarkeit und Zugriff

- Technische Vorkehrungen
- Vernetzung und Professionalisierung

Mit Hilfe dieser Empfehlungen können Leitlinien für die Anwendung in der Praxis erstellt werden, die dann über eine nachhaltige Koordinationsstruktur verbreitet werden können.

Vertrauenswürdige Langzeitarchivierung – Hauptbestandteile und Aufgaben

In Anlehnung an das OAIS-Referenzmodell besteht die vertrauenswürdige Langzeitarchivierung aus den folgenden Hauptbestandteilen.

- **Erstellung/ Erwerbung/ Beschaffung und Erschließung** (vgl. *Ingest* und Metadaten)
- **Bewahrung/ Speicherung/ Verarbeitung** (vgl. *Archival Storage* und *Data Management*)
- **Bereitstellung/ Nutzung** (vgl. *Administration*, *Data Management* und *Access*)
- **Erhaltung** (vgl. *Preservation Planning* sowie Migration und Emulation)
- **Zugriffsregelung** (vgl. *Access*)

Die Grundlage bzw. die Ausgangsposition ist auf folgende Punkte zurückzuführen:

- Vermehrter Einsatz digitaler Kommunikationssysteme
- Enorme Ansammlung von digitaler Information
- Unterschiedliche Nutzer
- Verschiedene Standards
- Unterschiedliche technische Modelle und Implementierungen
- Unterschiedliche wirtschaftliche Prinzipien
- Unterschiedliche Anforderungen

In Bezug auf die Umsetzung der vertrauenswürdigen Langzeitarchivierung lassen sich die Ziele wie folgt definieren:

- Kommunikation und ein Informationsaustausch zwischen allen Beteiligten
- Kooperation
- Metadatenaustausch zwischen digitalen Langzeitarchiven
- Austausch von digitalen Objekten
- Austausch von Archivierungstechnologien
- Standardisierter Austausch von Archivierungsinformation: Formate, Beschreibungen, Implementierungen

In Bezug auf vertrauenswürdige Bestandserhaltungskonzepte für digitale Ressourcen lassen sich zwei tragende Teilziele festhalten:

- **Substanzerhaltung** Die unversehrte (integere) und unverfälschte (authentische) Bewahrung der digitalen Information, d.h. die Substanzerhaltung der Dateninhalte.
- **Verfügbarkeit** Die langfristige Verfügbarkeit und Benutzbarkeit für bestimmte Benutzergruppen.

Zur Durchführung der *Substanzerhaltung* zählt in diesem Zusammenhang die Trennung der Inhalte von der Repräsentationsform – Auflösung der Abhängigkeit und Überführung in ein homogenes Speichersystem sowie die Übergabe an verantwortliche archivierende Institution(en) mit automatisierten Kontrollmechanismen, die den kontinuierlichen systeminternen Datentransfer überwachen. Da technische Plattformen einer kurzen Halbwertszeit unterliegen müssen Datenträger, -modelle, -schemata, -formate ständig gewechselt werden – Migration der Datenbestände. Würden Inhalte (Datensubstanz) nicht von dem Datenträger (Medium) getrennt werden, wäre eine dauerhafte

Substanzerhaltung unmöglich. Leider haben technische Maßnahmen zum Schutz der Verwertungsrechte (z.B. Kopierschutzverfahren) einen einschränkenden Einfluss, d.h. sie koppeln die Information an ein Medium.

Die Substanzerhaltung fungiert als Voraussetzung für die langfristige Gewährleistung der Verfügbarkeit und Benutzbarkeit digitaler Information. Im Kontext der *Verfügbarkeit*, der langfristigen Erhaltung der Benutzbarkeit und der Gewährleistung der Verfügbarkeit digitaler Information stößt man unweigerlich auf die Problematik, dass die Daten interpretiert werden müssen, um zugänglich zu sein, wie zuvor bereits beschrieben wurde. Dazu sind bestimmte technische Nutzungsumgebungen (Betriebssysteme, Anwendungsprogramme), auch Abspielsystem genannt, erforderlich. Aufgrund der zuvor genannten kurzen Halbwertszeit nicht nur technischer Plattformen, sondern auch technischer Nutzungsumgebungen, sind solche Umgebungen in der Regel langfristig überholt und nicht mehr verfügbar. Es bedarf einer Entwicklung von Strategien zur Lösung dieses Problems und damit für die Entwicklung vertrauenswürdiger Langzeitarchivierungssysteme. *Strategien* sind, wie in [BRSS03] detailliert beschrieben:

- **Migrationsverfahren** Transformation von einer technischen Umgebung in eine andere.
- **Emulationsverfahren** Lauffähiges Nachbilden von Systemumgebungen.

Während Migrationsverfahren in der Regel auf einfachere Datenstrukturen oder Generationswechsel von Datenträgertypen angewendet werden, sind komplexe digitale Objekte das typische Anwendungsfeld von Emulationsverfahren. Jedes dieser Verfahren bringt seine eigenen Vor- und Nachteile mit sich. So ist die Migration eine vergleichsweise etablierte Prozedur. Sie gestaltet die Nutzung der Objekte einfacher, weil hierzu kein zusätzliches Abspielsystem benötigt wird. Allerdings ist Migration ein fortwährender Prozess, welcher nur schwer kalkulierbar ist. Die Emulation verbucht für sich, dass zu dem Objekt selbst umfassend Zusatzangaben mit der Spezifikation seiner Umgebung archiviert werden. Umgekehrt ist das Verfahren mit hohem Aufwand verbunden und jede Nutzung des Objektes an die Emulationstechnologie gebunden ist.

Für die Vertrauenswürdigkeit von Archiven besteht die Problematik der Erhaltungsstrategien darin, dass zum einen die Migration oft in dem Aufheben von Originalität und Authentizität resultiert, d.h. bei einer Transformation wie der Migration besteht keine Garantie, dass diese für den Inhalt verlustfrei sind. Lediglich auf Bitebene ist dies nachweisbar. Die Erhaltung der Benutzbarkeit lässt sich also nicht immer mit der Erhaltung der ursprünglichen Ausprägung des „originalen“ Objektes vereinbaren. Zum anderen sind Emulationsverfahren oft sehr komplex und Aufwand und Kosten stehen in keinem vertretbaren Verhältnis zueinander. Für die Umsetzung von Strategien zur digitalen Langzeitarchivierung und der damit verbundenen Bestandserhaltung ist demnach eine Beschränkung auf die Kernfunktionen digitaler Ressourcen, den Informationsgehalt, notwendig. Es muss genau abgewogen werden, inwieweit es notwendig ist, die Authentizität und Integrität der zu archivierenden Information zu gewährleisten. Abgesehen von dem verbleibenden Bedrohungsrisiko der Information steht man bei der Realisierung digitaler Langzeitarchivierung unweigerlich vor dieser Entscheidung.

Langzeitarchivierungsmaßnahmen bzw. Strategien müssen demnach so konzipiert werden, dass trotz des Restrisikos Systeme angeboten werden, die vertrauenswürdig sind. Und „ein System ist dann vertrauenswürdig, wenn ein ausreichend glaubhafter und zuverlässiger Hinweis existiert, der zu glauben veranlasst, dass ein System bestimmten Anforderungen genügen wird.“ [Bis03] Laut Kriterienkatalog wird Vertrauenswürdigkeit (*Trustworthiness*) „als Eigenschaft eines Systems angesehen, gemäß seinen Zielen und Spezifikationen zu operieren (d.h. es tut genau das, was es zu tun vorgibt). Aus Sicht der IT-Sicherheit stellen Integrität, Authentizität, Vertraulichkeit und Verfügbarkeit Grundwerte dar. IT-Sicherheit ist somit ein wichtiger Baustein für vertrauenswürdige digitale Langzeitarchive.“ [Nes06]

Konzeption und Gliederung der Expertise

Die hier vorliegende Expertise „Vertrauenswürdige und abgesicherte Langzeitarchivierung multimedialer Inhalte“ soll beleuchten, welchen Bedrohungen digitale Langzeitarchive multimedialer

Inhalte potenziell ausgesetzt sind und wie diesen Bedrohungen durch Integration des Bausteins IT-Sicherheit in Strategien und Konzepten der digitalen Langzeitarchivierung begegnet werden kann.

Auf der Grundlage der Ergebnisse existierender Studien, namentlich der nestor-Expertise „Perspektiven der Langzeitarchivierung multimedialer Objekte“, der nestor-Expertise „Vergleich bestehender Archivsysteme“, dem nestor-Kriterienkatalog „Vertrauenswürdige Digitale Archive“ und der Studie „Integration von Archiv-Metadaten in einem zentralisierten Metadaten-Repository“ des Instituts für Rundfunktechnik (IRT) gibt **Kapitel 2** einleitend einen **allgemeinen Überblick über den Anwendungsbereich** der vertrauenswürdigen und abgesicherten Langzeitarchivierung. Weiterhin werden die **Anforderungen** an und **Annahmen** über vertrauenswürdige und abgesicherte Systeme zur Langzeitarchivierung digitaler Multimediainhalte aufgestellt. In diesem Zusammenhang werden zudem relevante Begriffsbestimmungen und Grundlagen dargelegt. Dazu zählen die Beschreibung von Medien, Medientypen, Formate und Formattypen sowie deren Wechselbeziehungen und die Unterscheidung zwischen Format- bzw. Medienwechsel und Format- bzw. Medienbrüchen ebenso wie Beschreibung und Erweiterung des Digitalen Objekts, die detaillierte Darlegung von Metadaten und deren Bedeutung sowie die kurze Zusammenfassung des OAIS-Referenzmodells. Bestandserhaltungsmaßnahmen wie die Migration oder Emulation werden hier nicht noch einmal aufgeführt, da sie in [BRSS03] bereits detailliert beschrieben sind. Bereits hier werden die Schwerpunkte und Inhalte der IT-Sicherheit mit einbezogen.

Zwei **Szenarien** dienen dieser Expertise als praktische Beispiele, anhand derer die Integration von IT-Sicherheit illustriert wird. **Kapitel 3** charakterisiert allgemein die exemplarischen Langzeitarchivierungssysteme wie sie zum einen in **Hochschul-Medienzentren** und zum anderen in **öffentlich-rechtliche Rundfunkanstalten** zum Einsatz kommen. Dabei werden wesentliche **Kenngrößen** für die exemplarischen Langzeitarchivierungssysteme, wie Menge, Art und Ort anfallender Daten, eingesetzte technische Systeme, Art und der Umfang der technischen Aufbereitung der Daten sowie die technische Infrastruktur erhoben. Insbesondere die Beschreibung der technischen Plattformen, deren Eigenschaften und Mechanismen zur Speicherung von hochkomplexen und verteilt liegenden Inhalten (Werkzeuge, Architekturen, Lösungen, Lösungen gegenüber hochgradig verteilte Lösungen) sowie die Inhaltsbeschreibungen und Rechte-Daten sowie Metadaten werden dabei betrachtet. Die Beschreibung der exemplarischen Langzeitarchivierungssysteme nennt ihre grundsätzlichen Komponenten der allgemeinen technischen Infrastruktur in Funktion und Struktur sowie die Informationsflüsse.

Darauf aufbauend gibt das **Kapitel 4 Ansatzpunkte für die Nutzung von Langzeitarchivierungstechniken** in den beiden Szenarien. Als Grundlage leistet diese Expertise für Hochschul-Medienzentren einerseits und für Rundfunkanstalten andererseits die **Abstraktion** der exemplarischen Langzeitarchivierungs-Einzelsysteme anhand des OAIS-Referenzmodells. Die Abstraktion unterscheidet Akteure im Umfeld des Archivs, Architektur und Rollen sowie Daten. Die Darstellung der Informationsflüsse betrachtet die Vorgänge Bestandserweiterung, Bestandsbereitstellung und Bestandserhaltung. In diesem Zusammenhang werden die speziellen Sicherheitsaspekte analysiert und die konkreten organisatorischen und technischen Anforderungen aufgestellt. Die strukturierte Abstraktion erlaubt einen Vergleich der beiden Szenarien in Bezug auf **Ähnlichkeiten und Unterschiede**.

Im **Kapitel 5** werden **Herangehensweisen zur Integration von Sicherheitstechnologien**, wie digitale Signaturen und Wasserzeichen, für eine vertrauenswürdige und abgesicherte Langzeitarchivierung multimedialer Inhalte erforscht. Zunächst werden dazu Sicherheitsaspekte allgemein dargestellt und die Auswirkungen ihrer Verletzung beschrieben. Anhand der Sicherheitsaspekte werden Soll-Anforderungen für eine vertrauenswürdige und abgesicherte Langzeitarchivierung multimedialer Inhalte aufgestellt und, basierend auf den existierenden Studien und in Verbindung mit der Erhebung der exemplarischen Langzeitarchivierungssysteme, dem Ist-Zustand systematisch gegenübergestellt. Eine **Validierung der praktischen Einsetzbarkeit** wird an Beispielen in den einzelnen Szenarien reflektiert, wobei auch potentielle Hürden der praktischen Umsetzung aufgezeigt werden.

Im **Kapitel 6** wird der **Handlungsbedarf** aufgezeigt, inwieweit Sicherheitstechnologien in Langzeitarchiven integriert werden sollten bzw. dies möglich ist. Darauf aufbauend wird versucht, einen möglichen **Standardisierungsbedarf** abzuleiten, in dem sich das Kompetenznetzwerk Langzeitarchivierung als Ganzes engagieren und **Aktivitäten initiieren** muss. Des Weiteren wird in diesem Zusammenhang der Einsatz von **potentiellen und neuartigen Technologien**, wie hochgradig verteilte Speichertechnologien evaluiert, um Potentiale für den Einsatz von **Grid- und anderen Virtualisierungstechnologien** aufzuzeigen.

Ein **Fazit** mit **Zusammenfassung** und **Ausblick** schließt im **Kapitel 7** diese Expertise ab.

2 Grundlagen, Anforderungen und Annahmen der vertrauenswürdigen und abgesicherten Langzeitarchivierung

In Kapitel 2 werden existierende Studien und deren Bedeutung für die vertrauenswürdige und abgesicherte Langzeitarchivierung zusammengefasst. Diese Studien bilden die Grundlage der vorliegenden Expertise. Aufbauend auf den Studien und Berichten sowie weiterführender Literatur werden im zweiten Abschnitt generelle Anforderungen und Annahmen an Systeme zur vertrauenswürdigen und abgesicherten Langzeitarchivierung zusammengetragen.

2.1 Existierende Studien

In diesem Abschnitt werden vorangegangene Studien aufgearbeitet. Diese Studien sind unterteilt in Studien, die innerhalb des nestor-Kompetenznetzwerks entstanden sind und Studien und Berichte des Instituts für Rundfunktechnik (IRT).

2.1.1 nestor-Studien

nestor-materialien 3: Vergleich bestehender Archivierungssysteme [Bor05]
Das Ziel dieser Studie war es, eine Basis für die Bewertung und Auswahl von Archivierungssystemen zu schaffen. Folgende Punkte zählen dabei zu den Inhalten:

- Zusammenfassung der Angebote
- Funktionalität
- Systemarchitektur (vgl. OAIS)
- Beschreibung Inhalt (digitale Objekte)
- Frage nach der Notwendigkeit spezieller Applikationsumgebungen
- Ingest-Verfahren
- Metadaten (technisch, inhaltlich, rechtlich, administrativ), deren Generierung sowie Standards
- Technologien
- Sicherheit: Gewährleistung Integrität und Authentizität der Daten
- Benutzerschnittstelle: Kosten/ Rechte
- Langzeitaspekt
- Schnittstellen zu anderen Systemen
- Dokumentation
- Kosten
- Vergleichskriterien

In dieser Studie wurde festgestellt, dass eine dynamische Entwicklung vorherrscht. Es existiert eine Vielzahl von Produkten, die verwendet werden. Die Anforderungen entstammen verschiedenen

Anwendungsbereichen und weisen unterschiedliche Lösungsansätze auf. Es wurde ein Kriterienkatalog für eine Produktbewertung aufgestellt, um so ein Rating und letztendlich ein Ranking durchzuführen. Es erfolgt eine ausschließliche Betrachtung „reiner" Archivierungssysteme, bei der die Archivierung an sich Kernaufgabe ist. Ausgeschlossen sind:

- Reine Entwicklungswerkzeuge
- Anwendungsneutrale Grundsysteme
- Produkte, bei denen die Archivierung nur die Teilfunktion einer umfassenden spezialisierten Anwendung darstellt

Bezüglich der Abdeckung des Langzeitaspekts, wurde festgestellt, dass bis auf wenige Ausnahmen, keine expliziten Mechanismen vorhanden oder Vorkehrungen getroffen sind, um die Inhalte auf lange Zeit zu erhalten und dass für Langzeitarchivierung anerkannte Techniken fehlen.

In Bezug auf die Archivierung, den Archivierungsbegriff und die Kriterien wurden folgende zwei Aspekte festgestellt:

- Es herrschen eine sehr breite Interpretation, unterschiedliche Begrifflichkeiten, divergierende Entwurfsziele mit wenig Bezug zu konzeptionellen Modellen wie OAIS vor.
- Probleme treten bei der Erstellung von Anforderungen und Systemen zur Archivierung auf, begründet durch:
  - Technologienahe Kriterien: Nennung der Lösungen ohne konzeptionelle Einordnung
  - Unterschiedliche Begrifflichkeiten
  - Mangelnde Allgemeingültigkeit
  - Vernachlässigung nicht-funktionaler Aspekte wie Aufwand für laufenden Betrieb oder Qualität des Produkte

In der Studie wurden die folgenden Aspekte bzgl. OAIS-Referenzmodell (*Open Archival Information System*) festgehalten:

- Das OAIS-Referenzmodell dient zur Erfassung der Kernfunktionalität eines Archivierungssystems.
- Das OAIS-Referenzmodell beschreibt eine als Archiv zu bezeichnende Organisation aus Menschen und Systemen, welche die Verantwortlichkeit übernommen hat, Information zu erhalten und sie für eine bestimmte Zielgruppe (*Designated Community*) verfügbar zu machen.
- Das OAIS-Referenzmodell ist keine Spezifikation eines Entwurfs oder Implementierung, tatsächliche Implementierungen können die Funktionalität anders gruppieren oder aufbrechen.
- Das OAIS-Referenzmodell beschreibt weiterhin Funktionalitäten, die sich nicht auf softwaremäßige Implementierungen beziehen, wie Bestandserhaltungsplanung (*Preservation Planning)*.
- Im OAIS-Referenzmodell fehlen Systemmerkmale, die für eine Bewertung einer konkreten Implementierung nötig sind (z.B. Kosten).

Bezüglich des Langzeitaspekts sind in der Studie weiterhin folgende Aussagen getroffen worden: Bisherige Methoden bzw. bisherige konkrete Datenmodelle zur Langzeitarchivierung lassen sich nicht aus OAIS ableiten. Eine Migration (Zugriff, Transformation, Rekonstruierung) des Inhalts und seiner Organisation erfolgt aus den bisherigen Systemen bzw. Systemstrukturen. Es wurden Standards und Produkte betrachtet, die den Langzeitarchivierungsaspekt nicht direkt ansprechen. Funktionale, auf Langzeitarchivierung bezogene, inhaltsorientierte Systemeigenschaften sind:

- Objekte, Objektorganisation
- Metadaten, Metadatenorganisation
- Organisation Objekte – Metadaten
- Sicherung der Integrität dieser Organisation

Nicht-funktionale Kriterien dagegen sind:

- Aufwand
- Qualität: Sicherstellung der Verfügbarkeit des Systems und insbesondere hinsichtlich Langzeitarchivierungsaspekten Sicherstellung der Verfügbarkeit der Inhalte
  - Funktionale Kriterien (Technik)
  - Migrierbarkeit der Inhalte in neue Systeme (Technik)
  - Hohe Verbreitung und Nutzerzahl verbunden mit der entsprechend wertvollen Datenmenge (Organisation)

Es ist eine Tendenz zur Modularisierung und zur logischen und physischen Verteilung der Systeme zu vermerken, um so einen ortsunabhängigen Zugang zu ermöglichen. Konzepte zur Unterstützung von Langzeitarchivierungsaspekten sind wenig vorhanden. Diese sind unter anderem:

- Dateiformatregistrierung
- Handlesysteme zur persistenten Identifikation
- Übernahme von Konventionen (Standards) im Bereich der Metadaten
- Konvertierung in bestimmte Ablageformate
- Migrationsunterstützung z.B. durch einfach zu interpretierende Exportformate
- Beschreibung des Kontextes der technischen Nutzung z.B. durch Archivierung der Softwareanteile der Abspielumgebung
- Universal Virtual Computer (UVC)

Die Überlebensfähigkeit eines Systems und seiner Inhalte ist abhängig von seinen Systemmerkmalen. So ist die Leistungsfähigkeit bisheriger Modelle unterschiedlich in Bezug auf die Flexibilität zur Organisation von Objekten und Metadaten und Festschreibung der Organisation. Ausgewiesene Schwachpunkte, die in der Studie benannt wurden sind:

- Flache oder starre Objekthierarchien
- Fehlende Definierbarkeit der Semantik der Objektbeziehungen
- Eingeschränkte Metadatenmodelle (fehlende Elemente zur Beschreibung von Repräsentationsinformation, eingeschränkte Zuordenbarkeit zu Objekten)
- Unkontrolliertes Vokabular für Metadatenelemente und deren Werte

nestor-materialien 5: Perspektiven der Langzeitarchivierung [Coy06]

Das Ziel dieser Studie war es, Empfehlungen für eine Langzeitarchivierung von digitalen multimedialen Artefakten zusammenzustellen. Folgende Punkte zählen dabei als *Empfehlungszusammenfassung* zu den Inhalten:

- Es existieren keine gesicherten Strategien für die langfristige (über 100 Jahre) Speicherung und Nutzung multimedialer, digitaler Daten.
- Förderung „proaktiver“ Erzeugung von Daten.
- Zukünftig sind keine dauerhaften Speichermedien vorhanden aufgrund der verkürzten Haltbarkeit.
- Bestandserhaltung digitaler Objekte als dynamischer Prozess der Speicherung.
- Kostenfaktor Umkopieren und Umkodieren zur Anpassung an neue Geräte und Systeme muss berücksichtigt werden.
- Pflege von Metadaten von entscheidender Bedeutung
- Technische und organisatorische Sicherung der Authentizität und Integrität digitaler Objekte ist notwendig.
- Vorzug von rechtlich fixierten und vollständig offen gelegten Standards.

- Geeignete Backup-Strategien zur Datensicherung wenn mögl. auf technisch verschiedene Speichermedien, auch analog
- Entwicklung von Emulatorprogrammen
- Trennung und Verteilung von Datensammlung und Datenspeicherung/ Datensicherung
- Bei verteilten Systemen und Netzzugriffen Speicherung bei vertrauenswürdigen Dritten
- Beachtung rechtlicher Regelungen

Anforderungen, Herausforderungen und Perspektiven:

- Perspektive besteht in der Bereitstellung der Materialien im Netz
- DVD und CD als Trägermedien sind nur Übergangslösungen

Der digitale Bestand:

- Ziel ist eine vollständige Migration der digitalen Materialien in den nicht mehr ortsgebundenen Netzzugriff – abhängig von den Aufgaben und Lizenzen als Intranet oder als Internetangebot.
- Unterscheidung zwischen selbsterstelltem und nach Vorgaben fremderstelltem Material.
  - *Selbsterstellte Materialien* sind in Bezug auf die künftige Verwendung nach entsprechenden Standards auf geeigneten Speichermedien mit präzisen Metadaten herstellbar und einsetzbar.
  - *Fremderstellte Materialien* brauchen übergreifende Austauschformate, wobei offene und freie Standards und eine gemeinsame, möglichst standardisierte, Metadatenstruktur eine entscheidende Rolle spielen.
- Es sollen keine Medientypen bei der Bestandserschließung, Bereitstellung und Sicherung ausgeschlossen sein/ werden.
- Organisation digitaler Materialien: Wie sind sie gespeichert? Bitströme
- Unterscheidung der digitalen Backup-Strategien: Lokal oder vernetzt mit qualitativ gleichwertigen, technischen Kopien.
- Sicherung der Integrität und Authentizität:
  - Unveränderter Erhalt des Sammelguts, Schutz vor illegalen Manipulationen, Sicherung der Urheberschaft und Unversehrtheit.
  - Es besteht eine hohe Gefahr der Manipulierbarkeit da spurlose Veränderungen einfach durchzuführen sind.
  - Entwicklung und Anwendung angemessener Bestandssicherungsstrategien.
- Rechtliche Verpflichtungen (Kopierschutz und Urheberschaft, Signaturgesetz) beachten und integrieren.
- Handhabbare und kostengünstige Durchsetzung.

Speichermedien und Speichermaterialien für multimediale Artefakte bezeichnen die eingesetzten materiellen Trägermedien. Diese gibt es in verschiedenen Ausprägungen:

- Papierspeicher und Ausdrucke: Unbrauchbar für multimediale Datenbestände.
- Microfilm und Computer Output on Microfilm: Sind als analoge Speicherlösungen nicht geeignet für multimediale Datenbestände, nur als spezielle Lösung für die Archivierung statischer Daten (Texte), kein digitaler Zugriff und keine Erschließung.
- Rosetta Stone: Nicht alle digitalen Formate können abgebildet werden. Bieten keine Sicherungsmöglichkeit für programmgestützte, interaktive, dynamische Medien.
- Optische Plattenspeicher wie CD-ROM, CD-R(W), Magnetplatten und Magnetbänder: Zählen zum Kernbestand der multimedialen Speichermedien, wobei deren Haltbarkeit problematisch sind.
- Iridium-CD-Datenspeicher: Keine praktischen Einsätze bisher.

- Magnetische Plattenspeicher: Dominierende Speichertechnik.
- Magnetbänder: Generell gut geeignet, eingeschränkt jedoch durch die sequentielle Zugriffstechnik.
- Direktzugriffsspeicher-Halbleiterspeicher (Flash-Speicher, MRAM): Potential für eine Verwendung besteht.
- Holografische Speicher: Ungewiss.
- Das Netz als Speicher: Zukunft mit regelmäßigem Umkopieren als notwendige Forderung der Bestandserhaltung.

„Leider bietet Umkopieren keine hinreichende Bestandserhaltungsstrategie. Trotz der recht beschränkten Lebenszeit technischer Digitalspeicher mit der Folge unlesbarer Medien scheint die Wahrscheinlichkeit größer, dass Daten durch technologische Alterung, also veraltete Geräte und Software, verloren gehen werden."

Formate:

- Technische Speicherformate, logische Formate, algorithmisierte Formate
- Metadaten
- Texte
- Daten, Datenbanken, Tabellenkalkulationsprogramme
- Programme
- Formatierte Texte (MS Word, RTF, TeX)
- PDF
- HTML, SGML, XML
- Schriften
- Grafik, CAD
- Fotos, Bilder
- Video
- Audio (Musiknotationen, Tonaufnahmen, gesprochene Sprache)
- Kompressionstechniken
- Spiele und anderes interaktives Material
- Multimedia (SMIL - Synchronized Multimedia Integration Language)
- Metadaten (Dublin Core – Semantische Minimaldaten) Metadatenformat
- Semantic Web

Emulation zur langfristigen Bereitstellung von komplexen Multimediadokumenten und Programmen:

- Langfristige Bereitstellung originaler Geräte und Software
- „Einfrieren" einer bestimmten technischen Entwicklungsstufe für spätere Nutzung gescheitert
- Bestandserhaltung durch Emulatoren
- Grenzen bei der Nachbildung proprietärer Betriebssysteme ohne Kenntnis des Quelltextes oder ohne rechtliche Nutzungsmöglichkeit evtl. vorhandener Programmtexte
- Emulation durch Einzug einer virtuellen Abstraktionsebene, wo ein Rechner in Softwareform angeboten wird, der selbst auf unterschiedlichen Geräte- und Betriebssystemplattformen implementiert ist, z.B. Java Virtual Machine (JVM), Loires Universal Virtual Machine (UVM)
- Problem: Emulatorkonzepte nicht standardisiert

Migrationspfade zur Bestandssicherung: Organisation, Zeithorizont, strategische Aspekte:

- Analoge Wandlungen zur Bestandserhaltung – analoges Speichern digitaler Dokumente
  - Computerausdruck keine Speicherung multimedialer Daten
  - COM (Computer Output on Microfilm): durchaus eine Variante von dauerhaftem Bestand, aber dennoch gänzlich ungeeignet für Multimedia-Bestandserhaltung, da interaktive Elemente zur Navigation oder zu anderen Zwecken mit analogem Material nicht oder nur mit völlig unverhältnismäßigem Aufwand realisiert bzw. simuliert werden können.
- Digitale Wandlungen analoger Objekte zur Bestandserhaltung
  - Komprimierte Speicherung digitaler Objekte
  - Einsetzbarkeit unter dem langfristigen Aspekt mehrfacher Umkodierung
  - Referenzen müssen frei zugänglich sein
- Digital-digitale Migrationen
  - Digitale Speichermedien sind technologisch und physikalisch gefährdet.
  - Strategie des Umkopierens
  - Verwendung stabiler, langfristiger verfügbarer Speicherformate, aber Veralterung verhältnismäßig schnell; es empfiehlt sich Umkodierung
  - Gefahr des unbemerkten Verlusts
  - Permanente und aktive Pflege des Bestands

„Realistisch gesehen wird der Migrationspfad hin zur digitalen Speicherung umfassend beschritten werden. Die Probleme der langfristigen Speicherung und Bereitstellung sind für Textdokumente und einfache Grafiken lösbar. Für multimediale Dokumente und alle Dokumente, die programmierte Aktivität verlangen, sind weitere Forschungs- und Entwicklungsarbeiten nötig – aber es gibt für multimediale Artefakte keine brauchbaren Alternativlösungen im nicht-digitalen Bereich.“

nestor materialien 8: Kriterienkatalog vertrauenswürdige digitale Langzeitarchive (nestor – Arbeitsgruppe „Vertrauenswürdige Archive –Zertifizierung“) [Nes06]

Das Ziel des Kriterienkatalogs war es, Anforderungen an vertrauenswürdige digitale Langzeitarchive festzulegen bezogen auf:

- A: Organisatorischer Rahmen
- B: Umgang mit Objekten
- C: Infrastruktur und Sicherheit

Der Abschnitt C „Infrastruktur und Sicherheit“ ist dabei sehr kurz gehalten.

Der Kriterienkatalog spiegelt ein offenes Vorgehen als Grundlage für eine Allgemeingültigkeit, eine Praxistauglichkeit und eine breite Akzeptanz der Ergebnisse wieder. Der Katalog dient zur Identifizierung von Kriterien für die Bewertung der Vertrauenswürdigkeit eines digitalen Langzeitarchivs in organisatorischer und technischer Hinsicht, dies in engem Kontakt mit unterschiedlichsten „Gedächtnisorganisationen“ und Produzenten, sowie weiteren Betroffenen und Experten. Es ist damit ein fundiertes, abgestimmtes und praxisgerechtes Hilfsmittel zur Erlangung und Darstellung von Vertrauenswürdigkeit im Kontext von Langzeitarchiven geschaffen. So kann der Nachweis der Vertrauenswürdigkeit durch eine Zertifizierung im Rahmen eines national und international standardisierten Verfahrens erlangt werden.

2.1.2 IRT-Studie und -Berichte

Das Institut für Rundfunktechnik (IRT) ist das zentrale Forschungs- und Entwicklungsinstitut bedeutender öffentlich-rechtlicher Rundfunkanstalten in der Bundesrepublik Deutschland, in Österreich und in der Schweiz.

Aufgrund seiner Position und Ausrichtung wurden Forschungsarbeiten und technische Entwicklungen des IRT als bedeutsame Quellen für diese Expertise betrachtet. Von den gesichteten Arbeiten des IRT wurden die Studie zur Vorbereitung der Strukturierung und Integration von Objekten der Rechte-

auskunft in einem verteilten Metadaten-Repository [NN01] sowie der Bericht zum Broadcast Metadata Exchange Format (BMF) zum Austausch von Metadaten [Ebn05] als unmittelbar relevant eingestuft und für diese Expertise herangezogen.

Studie zur Vorbereitung der Strukturierung und Integration von Objekten der Rechteauskunft in einem verteilten Metadaten-Repository [NN01]

Die Studie untersucht Anforderungen an und die technische Machbarkeit einer lose gekoppelten Anbindung der Archive von ARD-Mitgliedsanstalten an ein verteiltes Metadaten-Repository.

Die Erhebung betrachtete die Abteilungen Honorare-&-Lizenzen (HoLi) exemplarischer Anstalten des Verbundes. Die einzelnen Rundfunkanstalten betreiben eigene Datenbanken zur Verwaltung ihrer Rechteinformationen. Die systematische Erfassung der technischen Infrastrukturen zeigte auf, dass die zu Grunde liegende *technische Infrastruktur* variiert. Die systematische Erfassung der Mechanismen zur Repräsentation von Rechtedaten führte zu der Erkenntnis, dass die *Datenschemata* der Rundfunkanstalten sich formal voneinander unterscheiden und auch mit unterschiedlicher Bedeutung belegt sind.

Die sich in der Studie anschließende Machbarkeitsuntersuchung entwirft ein *verteiltes Rechercheystem*. Bei diesem Ansatz bleiben die bestehende Infrastruktur und die etablierten Arbeitsabläufe in den Rundfunkanstalten unverändert erhalten. Zusätzliche Komponenten übersetzen Rechercheanfragen und -ergebnisse einer übergreifenden Suche in die spezifischen Datenschemata der einbezogenen Rechtedatenbanken der Rundfunkanstalten. Das *Abbildungsverfahren* für die Kopplung der Rechteinformation berücksichtigt möglichst umfassend die Menge gemeinsam vorkommender Konzepte.

Bedeutung für diese Expertise erlangt diese Studie durch ihren Lösungsvorschlag für eine Kopplung eigenständiger Metadatenkataloge. Zudem gibt sie einen Einblick in übergreifend gemeinsame Konzepte von Rechtedaten.

Austausch von Metadaten – Broadcast Metadata exchange Format, BMF [Ebn05]

Der technische Bericht stellt Motivation, Anwendungsfälle, Anforderungen und Format des Datenmodells BMF zur Repräsentation von Metadaten in der Fernsehproduktion vor. Weil Metadaten für die langfristige Interpretation und Nutzung von archivierten Medienobjekten erforderlich sind ist der technische Bericht für diese Expertise relevant.

2.2 Generelle Anforderungen an und Annahmen über vertrauenswürdige und abgesicherte Langzeitarchivierungssysteme

Basierend auf der vorhergehenden Zusammenfassung und Aufarbeitung der bestehenden Studien und ihre Relevanz für die vertrauenswürdige und abgesicherte Langzeitarchivierung werden in diesem Abschnitt die generellen Anforderungen an und Annahmen über vertrauenswürdige und abgesicherte Langzeitarchivierungssysteme formuliert und dargestellt. In diesem Kontext werden Grundlagen in Form von Begriffsbestimmungen gegeben. Ausgangslage und Referenzmodell bildet der OAIS-Standard, der hier kurz umrissen wird verbunden mit der Beschreibung des Digitalen Objekts sowie von Metadaten und deren Bedeutung. Migration oder Emulation als Bestandserhaltungsmaßnahmen werden nicht aufgeführt, da sie in [BRSS03] bereits detailliert beschrieben sind und dort nachgelesen werden können. In der Beschreibung der Anforderungen und Annahmen wird bereits der Bezug zu den Schwerpunkten und Inhalten der IT-Sicherheit hergestellt.

2.2.1 Core Requirements für digitale Archive

Die Grundlage für Anforderungen an digitale Langzeitarchive bilden die Kernanforderungen (*Core Requirements for Digital Archives)* [DDNC07], festgelegt von Vertretern von vier Archivierungsorganisationen. Dort sind zehn grundlegende, ein digitales Langzeitarchiv charakterisierende Anforderungen wie folgt bestimmt:

1. Das Archiv verpflichtet sich, digitale Objekte für eine bestimmte Benutzergruppe (*Designated Community*) zu erhalten.
2. Das Archiv weist nach, dass es organisatorisch (Finanzen, Mitwirkende, Prozesse) befähigt ist, seinen Verpflichtungen (s. Punkt 1) verantwortlich gerecht zu werden.
3. Das Archiv erfasst, verwaltet und pflegt die erforderlichen Vertragsrechte und Rechtsgrundlagen und erfüllt Verantwortlichkeiten.
4. Das Archiv hat ein effektives und wirtschaftliches Grundsatzrahmenwerk.
5. Das Archiv erfasst und nimmt digitale Objekte auf (Ingest) gemäß ausgewiesenen Kriterien, die seinen Verpflichtungen und seinem Leistungsvermögen entsprechen.
6. Das Archiv erhält Integrität, Authentizität und Benutzbarkeit der von ihm über die Zeit aufbewahrten digitalen Objekte aufrecht.
7. Das Archiv erstellt, verwaltet und pflegt unerlässliche Metadaten sowohl über
   - Aktionen und Ereignisse auf digitalen Objekten während ihrer Aufbewahrung im Archiv als auch
   - den Entstehungsprozess, die Zugriffsunterstützung und den Nutzungskontext vor ihrer Archivierung
8. Das Archiv erfüllt die erforderlichen bzw. notwendigen Anforderungen an die Informationsverteilung (Dissemination).
9. Das Archiv hat ein strategisches Programm für die Planung der Erhaltung der Archivalien und deren Umsetzung (Preservation Planning)
10. Das Archiv verfügt über eine technische Infrastruktur, die adäquat ist, fortdauernd digitale Objekte zu erhalten und deren Sicherheit zu gewährleisten.

Diesen Anforderungen liegt die Annahme zu Grunde, dass für Archive jeglicher Art und Größe die Erhaltungsaktivitäten den Bedürfnissen und Verhältnissen der bestimmten Benutzergruppe(n) angepasst sind.

2.2.2 nestor-Kriterienkatalog

Die Ausgangslage für die Anforderungen ist festgelegt durch den nestor-Kriterienkatalog, wie in Abschnitt 2.1 bereits kurz beschrieben, in dem

a) der organisatorische Rahmen,
b) der Umgang mit den digitalen Objekten,
c) die Infrastruktur und Sicherheit

festgehalten sind. Abschnitt c. fällt verhältnismäßig kurz aus und besagt in den Punkten 13 und 14 mit:

> 13 Die IT-Infrastruktur ist angemessen.
>
> 13.1 Die IT-Infrastruktur setzt die Forderungen aus dem Umgang mit Objekten um.
>
> 13.2 Die IT-Infrastruktur setzt die Sicherheitsanforderungen des IT-Sicherheitskonzepts um
>
> 14 Die Infrastruktur gewährleistet den Schutz des digitalen Langzeitarchivs und seiner digitalen Objekte.

lediglich das „Was“ getan werden muss, um ein vertrauenswürdiges digitales Langzeitarchiv zu gestalten. Das „Wie“ bleibt hier offen. Hier bedarf es einer weitergehenden Spezifizierung. Genau an diesem Punkt setzt diese Expertise an. Denn über die IT-Infrastruktur hinaus müssen auch Aspekte, wie der Informationsfluss und die verwendeten Komponenten und Umgebungen betrachtet werden. Innerhalb dieser Expertise soll nun zusätzlich das „Wie“ erörtert werden. Wie sieht solch eine IT-Infrastruktur aus, die den Schutz des digitalen Langzeitarchivs und seiner digitalen Objekte gewährleistet, dabei die Sicherheitsanforderungen und Forderungen im Umgang mit Objekten umsetzt und angemessen aus wirtschaftlicher Sicht ist. Diese Expertise soll Antwort auf diese Frage geben.

IT-Sicherheit legt mit ihren Sicherheitsaspekten Integrität, Authentizität, Vertraulichkeit, Verfügbarkeit und Nachweisbarkeit/ Nicht-Abstreitbarkeit zunächst weiterführende Anforderungen fest, nämlich Anforderungen, die in den Punkten 6 und 10 der *Core Requirements for Digital Archives* explizit aufgelistet sind, darüber hinaus aber jeden Punkt betreffen.

2.2.3 Merksätze mit Anforderungen

In Anlehnung an *VOI (Verband Organisations- und Informationssysteme)* [KaRo97] sind die folgenden zehn Merksätze, die gleichzeitig Anforderungen beinhalten aufgelistet:

1. Jede Datei muss unveränderbar archiviert werden.
2. Es darf keine Datei auf dem Weg ins Archiv oder im Archiv verloren gehen.
3. Jede Datei muss mit geeigneten Retrievaltechniken wieder auffindbar sein.
4. Es muss genau die Datei wieder gefunden werden, die gesucht worden ist – gesuchte und gefundene Datei müssen identisch sein.
5. Keine Datei darf während ihrer vorgesehenen Lebenszeit zerstört werden.
6. Jede Datei muss in genau der gleichen Form, wie sie erfasst wurde, wieder angezeigt werden können.
7. Jede Datei muss zeitnah wieder gefunden werden können.
8. Alle Aktionen im Archiv, die Veränderungen in der Organisation und Struktur bewirken, sind derart zu protokollieren, dass die Wiederherstellung des ursprünglichen Zustandes möglich ist.
9. Elektronische Archive sind so auszulegen, dass eine Migration auf neue Plattformen, Medien, Softwareversionen und Komponenten ohne Informationsverlust möglich ist.
10. Das System muss dem Anwender die Möglichkeit bieten, die gesetzlichen Bestimmungen (BDSG, HGB/AO, etc.) sowie die betrieblichen Bestimmungen des Anwenders hinsichtlich Datensicherheit und Datenschutz über die Lebensdauer des Archivs sicherzustellen.

Um diesen Anforderungen gerecht zu werden, bestehen Archivsysteme im Allgemeinen aus Datenbanken, Archivsoftware und Speichersystemen, wie sie von verschiedenen Herstellern angeboten werden. Die Lösung besteht in einer so genannten Referenz-Datenbank-Architektur. Mittels einer *Referenzdatenbank* mit den Verwaltungs- und Indexkriterien wird auf einen Speicher referenziert, welcher die digitalen Objekte trägt. So können Informationen in separate Archivspeicher ausgelagert werden, unabhängig von den schnellen jedoch kostenintensiveren Online- Cachespeichern. Die Datenbank ist in der Lage, mittels eines Index ein Dokument im Speicher wieder zu finden und dem Konsumenten mit einem geeigneten Abspielsystem bereitzustellen.

2.2.4 Technische Aspekte

In Bezug auf Gestaltung und Bereitstellung von Systemen für eine vertrauenswürdige digitale Langzeitarchivierung spielt im Allgemeinen eine Reihe von *technischen Aspekten* eine ausschlaggebende Rolle:

- Speichermedium für die digitale Langzeitarchivierung
- Speichersysteme
- IT-Infrastruktur
- Verbindungen/ Kommunikation
- Darstellungsanwendung (Abspielsystem) und dessen Anbindung
- Prozess Einstellung ins Archivsystem (Ingest)
- Prozess Bewahrung der Inhalte und Metadaten (Planung, Migration, Emulation)
- Prozess Auslieferung an bestimmte Benutzergruppe (Access)
- Datenbanken
- Metadaten
- Vertrauensverhältnisse

2.2.5 Zentral organisierte Speichersysteme

Zentral organisierte Speichersysteme in der elektronischen Datenverarbeitung weisen klassischerweise einen *hierarchischen Aufbau* auf [WG06]. Es lassen sich grob vier Ebenen unterscheiden:

- **Online** Unmittelbar von den produktiven Geräten zugegriffener Speicher für die Materialbearbeitung. Online-Speicher zeichnet sich durch hohen Durchsatz und kleine Latenz aus. In der Regel realisiert durch Festplattenspeichersysteme.
- **Nearline** Nachgeschalteter Speicher, von welchem kurzfristig Material in den Online-Speicher überführt werden kann.
- **Archiv** Speicher mit hoher Kapazität, welcher Material vom Nearline-System aufnimmt, sobald sich dessen Kapazität erschöpft. Typischerweise realisiert mittels Bandrobotiksystemen.
- **Offline** Aus den Laufwerken und dem automatisierten Zugriff ausgelagerte Datenträger des Archivs.

In dieser Hierarchie (Abbildung 1) nehmen der *Preis* als auch die *Einfachheit der Handhabung* von oben nach unten ab. Genau entgegen gesetzt entwickeln sich *Kapazität* und *Latenz* der Speicherebenen.

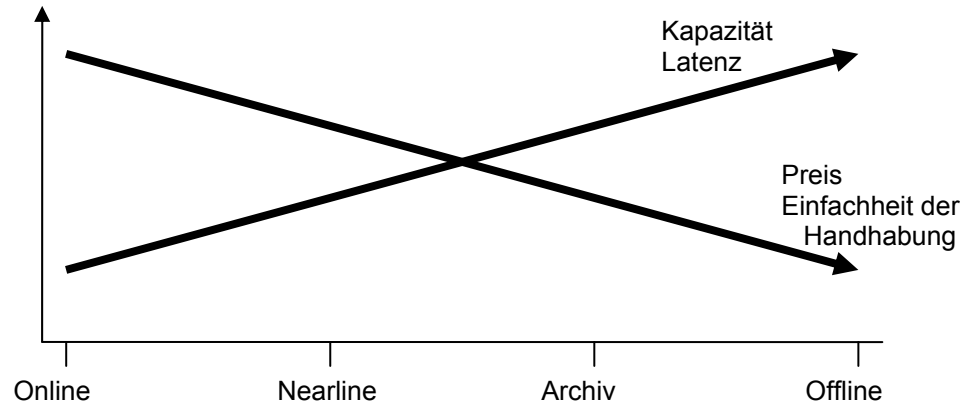

Abbildung 1: Hierarchischer Aufbau zentral organisierter Speichersysteme und deren Eigenschaften, vgl. [WG06].

2.2.6 OAIS-Referenzmodell

Um die Vertrauenswürdigkeit digitaler Langezeitarchive zu gewährleisten, werden allgemein akzeptierte Leistungskriterien für vertrauenswürdige digitale Archive versucht aufzustellen. Zu diesen Kriterien zählen zum einen die Konformität zum OAIS-Referenzmodell und zum anderen die Beständigkeit der institutionellen Struktur, von der das Archiv betrieben wird. Der ISO-Standard 14721:2001 Reference Model for an *Open Archival Information System* (OAIS) beschreibt die Infrastruktur eines digitalen Archivs in Form eines Modells. Kernfunktionen, welche die Abgrenzung und eindeutige Benennung von Funktionseinheiten, Schnittstellen und Typen von Informationsobjekten beinhalten sind im OAIS festgelegt. So konnte eine über die Grenzen der Anwendergemeinschaften Archive, Datenzentren und Bibliotheken hinweg geltende allgemeine Sicht auf ein digitales Archiv geschaffen werden.

Das OAIS-Referenzmodell ist generell eine ausschließlich logisch-konzeptuelle Beschreibung bzw. Definition – ein *Rahmenwerk*. Die tatsächliche Realisierung wird völlig offen gelassen. Es dient zur Zusammenfassung von Anforderungen, derer es bedarf, um Information für lange Zeit zu archivieren und einer bestimmten Nutzergruppe (*Designated Community*) und einem Anwendungsgebiet (*Application Domain*) zugänglich zu machen.[1] Im Zusammenhang der Evaluation der Anwendungsmöglichkeiten von IT-Sicherheit für eine vertrauenswürdige Langzeitarchivierung beschränkt sich diese Expertise in der nachfolgenden Darstellung auf die Zusammenfassung einiger wesentlicher

[1] Die komplette Terminologie kann im „Blue Book“ [CCSDS02] nachgelesen werden.

Aspekte der OAIS-Konzeption. Neben der Darstellung der Archivumgebung, Akteure und Aufgaben bzw. Prozesse und der Definition von Information Package betrifft dies auch die Handhabung und Abgrenzung von Daten und Information.

Umgebung der Langzeitarchivierung und Akteure

Die allgemeine Umgebung der Langzeitarchivierung laut OAIS ist in Abbildung 2 dargestellt. Das Archiv wird von Produzenten, Nutzern/ Konsumenten und dem Management beeinflusst. Der *Produzent* stellt die Information bereit bzw. fügt sie dem Archivsystem hinzu. Er spielt die entscheidende Rolle in der Erwerbung/ Beschaffung, kann aber je nach Archivsystem auch im Rahmen der Erschließung agieren. Der *Konsument* ist im Kontext der Bereitstellung aktiv. Er interagiert mit dem Archivsystem und fragt für ihn interessante Information ab. Dass diese Information auf Abruf verfügbar und benutzbar ist, dafür muss das Archiv sorgen. Das *Management* betrifft alle Bestandteile eines Archivsystems: Erwerbung/ Beschaffung/ Erschließung, Bewahrung/ Verarbeitung, Erhaltung sowie Bereitstellung/ Nutzung mitsamt Zugriffsregelung. Es legt die globalen Richtlinien fest, ist jedoch nicht in die alltäglichen Archivoperationen involviert. Es bestimmt lediglich, welche Information gesammelt werden soll und welche ausgeschlossen ist.

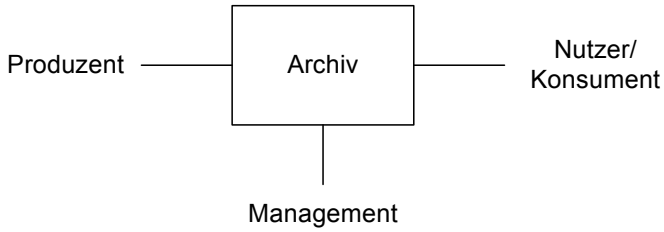

Abbildung 2: Umgebung der Langzeitarchivierung gemäß OAIS-Referenzmodell [CCSDS02].

Grundstruktur, Akteure, funktionale Einheiten und Aufgaben

In Abbildung 3 ist das Langzeitarchivsystem mit seinen funktionalen Entitäten, den Akteuren in seiner Umgebung mitsamt den hauptsächlichen Informationsflüssen dargestellt.

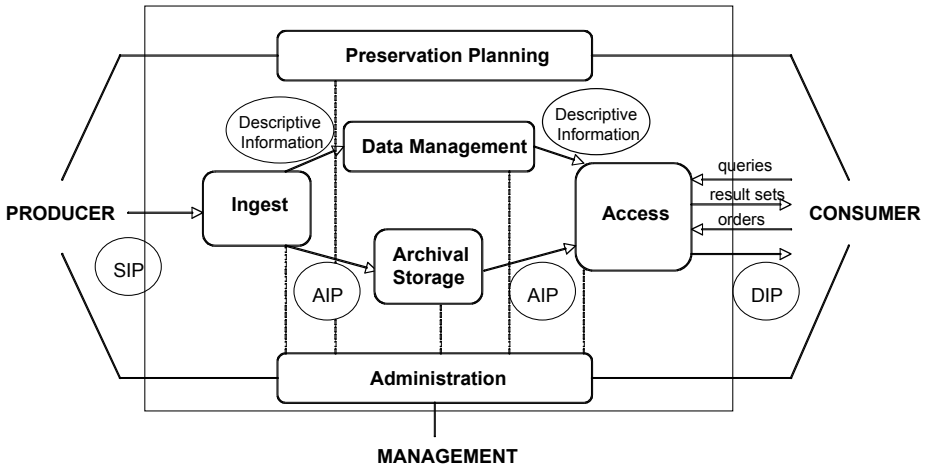

Abbildung 3: Funktionale Entitäten gemäß OAIS [CCSDS02].

Das OAIS beschreibt mehrere *Funktionseinheiten*, die dem Datenfluss und den Arbeitsabläufen des Archivs entsprechend angeordnet sind: Eingangsbearbeitung (*Ingest*), Metadatenverwaltung (*Data Management*), Archivspeicherung (*Archival Storage*), Bestandserhaltungsplanung (*Preservation Planning*), Administration und Bereitstellung (*Access*). Tabelle 1 listet die Entitäten mit ihren Aufgaben und Aktivitäten. Die Entitäten werden wiederum in Teileinheiten untergliedert. So besteht die Funktionsentität zur Erhaltung der Langzeitverfügbarkeit wiederum aus vier Teileinheiten, deren

Aufgabe es ist, die Umgebungsbedingungen des Archivsystems zu beobachten, Auswirkungen von technischen Veränderungen möglichst früh zu erkennen und die Planungsgrundlage für die Erhaltung der Langzeitverfügbarkeit der im System gespeicherten Objekte zu liefern.

Tabelle 1: Entitäten im Archiv und Aufgaben.

Eingangsbearbeitung (Ingest)	Entgegennahme von zu archivierender Information und Aufbereitung der Information für Speicherung und Verwaltung innerhalb des Archivs. ▪ Aushandeln von Vereinbarungen mit Produzenten ▪ Entgegennahme von SIPs ▪ Kontrolle der Vollständigkeit und Authentizität des SIP ▪ Umwandlung SIP zu AIP ▪ Erzeugen von Descriptive Information (Metadaten) zu AIPs ▪ Weitergabe AIP an Archival-Storage-Prozess
Archivspeicherung (Archival Storage)	Speicherung und Erhaltung der Bitströme der archivierten Information. ▪ Übernahme AIPs von Ingest ▪ Verteilung AIPs auf verschiedene Speichermedien ▪ Speicherverwaltung ▪ Prüfung der Unversehrtheit der AIPs gegenüber Speicher-fehlern und periodisches Wiederauffrischen der Speichermedien ▪ Sicherstellung der Rekonstruierbarkeit der AIPs nach Systemausfall ▪ Auf Anfrage Weitergabe der AIPs an Access-Prozess
Metadatenverwaltung (Data Management)	Verwaltung von Descriptive Information (Metadaten) und von den Daten, die für das Funktionieren des Systems verantwortlich sind. ▪ Administration und Fortschreiben einer Datenbank, in der die Daten gehalten werden ▪ Verteilung AIPs auf verschiedene Speichermedien ▪ Durchführung von Anfragen an Archivdatenbank ▪ Aufbereiten der gelieferten Daten
Administration (Administration)	Sicherstellen des routinemäßigen Funktionieren des gesamten Archivs. ▪ Aushandeln von Bedingungen mit Produzenten, unter denen sie ihre Information an das Archiv sendet ▪ Kontrolle der Übereinstimmung der gelieferten SIPs mit den Standards des Archivs ▪ Verantwortlich für Hardware- und Softwarearchitektur des Archivs ▪ Überwachung von Systemfunktionen ▪ Untersuchung von Möglichkeiten zur Optimierung von System-funktionen ▪ Entscheidung über Fortschreiben bzw. Migration von Archivinhalten ▪ Entwicklung und Überwachung der für das Archiv verbindlichen Standards
Bestandserhaltungsplanung (Preservation Planning)	Sicherstellung des zukünftigen technischen Zugriffs auf die im Archiv gespeicherten Information ▪ Überwachung der Entwicklungen des Hard- und Softwaremarktes ▪ Prüfung der Lauffähigkeit der gespeicherten Information auf neu ent-wickelten Systemen ▪ Entscheidung über zu verfolgende Erhaltungsstrategie (Migration/

	Emulation)
Bereitstellung (Access)	Bereitstellung von Archivinformation gegenüber Konsumenten. ▪ Auf Konsumentenanfrage Durchsuchen des Archivinhalts ▪ Auf Konsumentenanfrage Erzeugen eines DIP ▪ Überwachung der Auslieferung des DIP an Konsumenten

Information Package (OAIS)

Im Kontext der digitalen Langzeitarchivierung geht es um die Erhaltung von Informationsinhalten, wozu das OAIS-Referenzmodell ein Rahmenwerk liefert, und nicht notwendigerweise um die Erhaltung der Repräsentationsform der Information. Dies ist durch die schnelle Entwicklung digitaler Technologien zu begründen. Welche Information im Zusammenhang mit der vertrauenswürdigen Langzeitarchivierung nötig ist und wie deren strukturelle Relationen untereinander ist, ist mit der Information Package Definition beschrieben. Das OAIS-Referenzmodell benutzt diese Definition, um Informationseinheiten innerhalb eines Archivs zu modellieren.

Die Aufgaben bzw. Prozesse orientieren sich dabei an so genannten Information Packages. Generell werden im OAIS-Referenzmodell drei Arten von Information Packages unterschieden:

- **Submission Information Package (SIP)** wird vom Produzenten ans Archiv gesandt.
- **Archive Information Package (AIP)** wird im Archiv gespeichert.
- **Dissemination Information Package (DIP)** wird an Konsumenten ausgeliefert.

Ein *Information Package* ist laut OAIS festgelegt als bestehend aus den beiden Hauptkomponenten *Content Information* und *Preservation Description Information (PDI)*, wobei es diese zu einer logischen Einheit vereint.

Die *Content Information* ist das eigentliche und vom Archiv zu bewahrende Informationsobjekt. Die Content Information beinhaltet das Information Object einschließlich jeglicher dazugehöriger Information. *Preservation Description Information* bezeichnet alle, zur angemessenen Bewahrung der Content Information notwendige Information in einem Archiv. Dies ist Information, welche die authentische Originalität und den Ursprung der zu archivierenden Information garantieren und dessen Beziehungen zu anderen Objekten im Archiv beschreiben.

Was fehlt sind die Beschreibungen von Beziehungen von Archivobjekten zu Objekten außerhalb des Archivs, was in einer vernetzten digitalen Umgebung besonders wichtig ist. Denn ein digitales Langzeitarchiv ist kein in sich abgeschlossenes System und Verbindungen nach außen und zu anderen Archiven und deren Objekten sind ständig vorhanden. Hier können beispielsweise Hypermedia-dokumente genannt werden, die mittels Verweise zu anderen Objekten außerhalb des Archivs verlinkt sind. Dieser Punkt muss in Zukunft verstärkt betrachtet werden, auch insbesondere mit Blick auf die Entwicklung vertrauenswürdiger und abgesicherter Langzeitarchive.

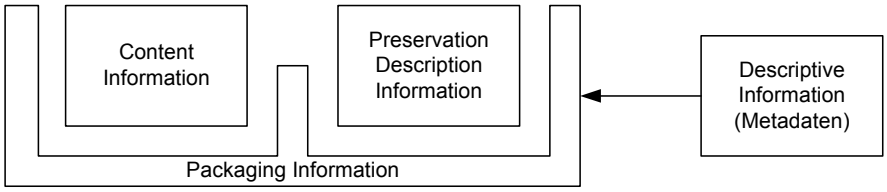

Abbildung 4: Information Package [CCSDS02].

Zur Verbindung der beiden Komponenten des Information Packages auf einem physikalischen Medium stellt das OAIS-Referenzmodell die so genannte *Package Information* zur Verfügung (Abbildung 4). Hier werden Beziehungen der beiden untergeordneten Informationspakete beschrieben, wie z.B. die Verzeichnisstruktur einer CD-ROM. Die *Descriptive Information* (Metadaten) beinhaltet

Informationen über die gespeicherten Information Packages und dessen Inhalte. So können die Packages im Archiv aufgefunden werden.

Daten und Information

Im OAIS-Referenzmodell wird grundsätzlich zwischen Daten (*Data Object*) und Information (*Information Object*) unterschieden. Das dort verwendete Informationsmodell basiert auf dem Verständnis nach Kuhlen.[2]

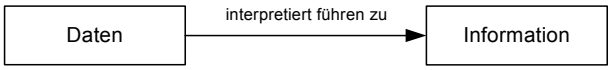

Abbildung 5: Daten und Information.

Information entsteht durch kontextorientierte Interpretation von Daten (vgl. Abbildung 5). *Daten* sind Objekte (analog oder digital) ohne jegliche Bedeutung. Daten können erst durch die Interpretation Bedeutung erhalten. Dann spricht man von Information. Dazu ist eine so genannte Wissensbasis (*Knowledge Base*) notwendig. Die Wissensbasis beinhaltet das nötige Wissen, um Daten zu verstehen. Ist eine Wissensbasis nicht vorhanden, können die Daten nicht verstanden werden. So liegt beispielsweise ein Text in englischer Sprache als Datenobjekt vor. Die Wissensbasis beinhaltet das Wissen über die englische Sprache und daher kann der Text nur verstanden werden, wenn auf eine solche Wissensbasis zurückgegriffen werden kann. Ansonsten ist zusätzliche Information zur Darstellung, die so genannte *Representation Information* notwendig, um in Kombination mit der vorhandenen Wissensbasis die Daten zu verstehen. Solche Repräsentationsinformation ist in diesem Fall ein Wörterbuch der englischen Sprache.

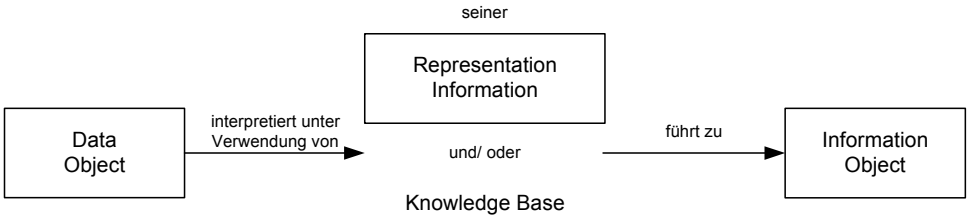

Abbildung 6: Zusammensetzung Information Object.

In Bezug auf digitale *Datenobjekte* beinhaltet die Repräsentationsinformation unterschiedliche Information, wie Information über das Format, die Art und Weise und der Ort des gespeicherten Objektes sowie Information, die dem Verständnis des eigentlichen Inhalts dienen. Für eine vertrauenswürdige Langzeitarchivierung bedeutet dies: Die Wissensbasis der Nutzergruppe und Anwendungsgebiet muss bekannt sein, um das Minimum an notwendiger Repräsentationsinformation für ein Verständnis des *Datenobjekte* ermitteln und anbieten zu können. Ein Informationsobjekt setzt sich aus dem Datenobjekt, der Repräsentationsinformation sowie der Wissensbasis zusammen, wie in Abbildung 6 skizziert. Dies ist skalierbar und gilt für alle Arten von Information.

2.2.7 Digitales Objekt

Parallel zum OAIS-Referenzmodell gibt es eine Beschreibung des digitalen Objekts [Thi02], die das digitale Objekt als Zusammensetzung aus drei *Ebenen* behandelt – der physikalischen Ebene, der logischen Ebene und der konzeptuellen Ebene. Diese Beschreibung ist für die Langzeitarchivierung digitaler Information besonders wichtig, da die Erhaltung der Information Mechanismen einschließt, die je nach Anforderung des Langzeitarchivierungssystems und je nach Art der Information eine Veränderung der Information und somit Medienbrüche auf unterschiedlichen Ebenen verursachen.

Eine Migration betrifft demnach eine andere Ebene als eine Emulation oder das einfache Kopieren. So wird beim Kopieren das Speichermedium gewechselt, was eine Änderung auf physikalischer Ebene bewirkt, nicht aber notwendigerweise auf der logischen Ebene oder der konzeptuellen Ebene, denn das

[2] http://www.inf-wiss.uni-konstanz.de/People/RK/publikationen.html

Format der Speicherung kann ebenso wie die Darstellung für den Menschen trotz einer Änderung auf physikalischer Ebene unverändert bleiben. Dies ist z.B. der Fall beim einfachen Kopieren eines Objektes von einem Datenträger auf einen anderen. Eine Migration von einem Datenformat in ein anderes bewirkt zunächst eine Veränderung auf logischer Ebene dahingehend, dass der Bitstrom (*Bitstream*) verändert wird. Dies kann sich auf die physikalische Ebene auswirken ebenso wie auf die konzeptuelle Ebene. Ist z.B. ein Objekt in ein Format geändert worden, welches eine Kompression einschließt, so kann dies zu Qualitätseinbußen führen, die wahrnehmbar sind. Eine Emulation wird dann eingesetzt, wenn die originale Information unverändert erhalten bleiben soll und muss.

Tabelle 2: Ebenen eines digitalen Objekts [Thi02].

Physikalische Ebene	Datenträger und physische Speichereigenschaften
Logische Ebene	Darstellungsform(at), Interpretation nach einer bestimmten Logik durch eine Anwendung – individuelles Datenformat einschließlich Bitstrom
Konzeptionelle Ebene	Eigentliche Darstellung des Signals/ Inhalts für den Anwender einschließlich des Informationsverarbeitungsprozesses des Anwenders

Die Erhaltung von digitalen Objekten ist mit verschiedenen Problemen verbunden. Es gibt wie bereits erwähnt grundsätzlich unterschiedliche Arten von Information, die eigentlich in ihrer ursprünglichen Form erhalten werden sollen. Da dies in digitalen Systemen nicht immer gänzlich möglich ist, soll zumindest eine bestimmte, durch Anforderungen festgelegte, Charakteristik einer Information unverändert erhalten werden. Das bedeutet, dass nicht zwangsläufig alle originalen Attribute einer Information in ihrem originalen Format erhalten werden müssen. Wichtig ist, dass eine Information in ihrer Substanz erhalten wird, gerade wenn sie von einer Plattform auf andere übertragen wird.

Thibodeau [Thi02] definiert ein *digitales Objekt* als ein Informationsobjekt jeglicher Art von Information unterschiedlichen Formats, das digital dargestellt ist. Alle digitalen Objekte sind Entitäten mit Mehrfachvererbung. Ein digitales Objekt lässt sich in drei Ebenen aufgliedern: physikalische Ebene, logische Ebene und konzeptuelle (kognitive) Ebene. Eine Kurzübersicht der Bedeutung der einzelnen Ebenen ist in Tabelle 2 aufgeführt.

Die **physische Ebene** des digitalen Objekts ist unabhängig von der Bedeutung bzw. Substanz und beinhaltet weder Syntax noch Semantik noch Morphologie der Information, sondern ausschließlich die *Zeichen*, die auf einem Informationsträger (Medium) eingeschrieben sind und dessen physische Speichereigenschaften. Entsprechend der Beschaffenheit des physischen Mediums wie DVD, CD-ROM, Festplatte, Blue-ray Disc, etc. ist die physische Eigenschaft der Zeichen unterschiedlich. Auf einer CD-ROM sind es die so genannten *Pits* and *Lands* während für magnetische Datenträger die Übergänge zwischen magnetisierten und nicht magnetisierten Teilchen charakteristisch sind.

Die **logische Ebene** des digitalen Objekts beinhaltet die Bits, die vom Datenträger gelesen und durch eine bestimmte Darstellungssoftware (Abspielsystem) nach einer bestimmten *Logik* interpretiert werden. In dieser Ebene ist es der Bitstrom, der das individuelle Datenformat des digitalen Objekts bestimmt und für das digitale Objekt charakteristisch ist. In der logischen Ebene ist festgelegt, wie die Information in Bits kodiert ist, wie verschiedene Kodierungen von einem in andere Formate transformiert werden und wie der eingehende Bitstrom transformiert wird zur Speicherung als physisches Objekt im Systemspeicher und der Output generiert wird für die Wiedergabe.

Die **konzeptuelle Ebene** des digitalen Objekts ist bestimmt durch die Information (Zeichen, Signale, Objekte der realen Welt) die vom Menschen oder einer Computeranwendung wahrgenommen wird. Die konzeptuelle Ebene eines digitalen Objekts ist die eigentliche *Information*, der Inhalt bzw. die Substanz, dem/ der durch die Wahrnehmung eine Bedeutung zugesprochen wird. Diese Ebene wird auch als kognitive Ebene bezeichnet. Kognition gleich Wahrnehmung. Der Inhalt selbst und die Struktur müssen in der logischen Ebene enthalten sein. Es gibt verschiedene logische Darstellungsformen für dasselbe konzeptuelle Objekt.

Im Zusammenhang mit digitalen Langzeitarchivierungssystemen ist diese Einteilung dahingehend von Vorteil, dass angepasste Erhaltungsmechanismen entwickelt werden können, um ein digitales Objekt den spezifischen Anforderungen angepasst auf lange Zeit zu sichern.

Strategien zur Langzeiterhaltung müssen Methoden beinhalten, welche die Datenintegrität über die Art der Speicherung hinaus sicherstellen. Solche Methoden können etwa die Migration (Kopieren), das Aktualisieren, das Verschieben, das Ausliefern und die Erneuerung/ Aktualisierung sein. Diese Methoden bewirken in der Regel eine Veränderung auf der physischen Ebene des digitalen Objekts, nicht aber notwendigerweise auf der logischen oder konzeptuellen Ebene.

Eine Komposition verschiedener digitaler Objekte, wie sie in digitalen Langzeitarchivierungssystemen oft vorkommt, geschieht auf logischer Ebene, wobei nicht die physische Eigenschaft des Speichermediums von Belang ist sondern der angemessene Ort zur Wiederauffindbarkeit. So sollte in einem digitalen Langzeitarchiv jedes eingehende digitale Objekt einen **persistenten Identifikator (*Persistent Identifier, PI*)** erhalten, um so im Archiv langfristig zu jedem Zeitpunkt wieder auffindbar und verfügbar zu sein. Auf diese Weise kann die Authentizität eines digitalen Archivobjektes gesichert werden.

Zur Langzeiterhaltung digitaler Objekte müssen die Anforderungen und Vorgaben für die zutreffende und korrekte Verarbeitung jedes Datentyps von Objekten dem System bekannt sein, und die entsprechende Software muss zur Verarbeitung bereitstehen. Dies bezieht sich auf die Verfügbarkeit der digitalen Archivobjekte, die auf diese Weise sichergestellt werden kann.

Beziehungen der Ebenen

Alle drei Ebenen beeinflussen sich gegenseitig, dies aber auf unterschiedliche Art und Weise. So sind beispielsweise zwei Objekte auf logischer Ebene identisch, aber aufgrund der Charakteristik ihrer unterschiedlichen physischen Speichermedien sind sie auf physischer Ebene verschieden. Ebenso kann ein Objekt, dass auf konzeptueller Ebene als gleich wahrgenommen wird, auf logischer Ebene und auch auf physischer Ebene völlig unterschiedlich sein. So kann ein Bild in unterschiedlichen Formaten auf unterschiedlichen physischen Speichermedien vorliegen und gleichzeitig als das gleiche Bild wahrgenommen werden. Die Ebenen können also in bestimmten Beziehungen zueinander stehen, was wie folgt zu beschreiben ist:

1 : 1 Ein digitales konzeptuelles Objekt besteht aus einer Datei und ist in einem bestimmten logischen Format an einem bestimmten physischen Ort auf einem Informationsträger gespeichert.

1 : n Ein digitales konzeptuelles Objekt besteht aus mehreren Dateien (logischen Einheiten), die an unterschiedlichen physischen Orten auf einem Informationsträger gespeichert sein können.

n : 1 Mehrere bzw. verschiedene digitale konzeptuelle Objekte können aus einer Datei (logischen Einheit) erzeugt werden wie z.B. bei einer Anfrage an eine Datenbank, in der ein Objekt logisch in einem bestimmten Format gespeichert ist und beispielsweise je nach Rechteverfügung in unterschiedlicher Form ausgeliefert wird.

Mehrere logische Objekte können auch zu einem physischen Objekt zusammengefasst werden.

n : m Mehrere bzw. verschiedene digitale konzeptuelle Objekte können aus unterschiedlichen Dateien (logischen Einheiten) erzeugt werden, die beispielsweise in der Datenbank verteilt gespeichert und ggf. an unterschiedlichen physischen Orten auf einem Informationsträger gespeichert sind.

Beziehung konzeptuelle Ebene zu physischer Ebene

Hier ist festzuhalten, dass dieses Modell lediglich ein digitales Objekt selbst behandelt, also nur auf Datenebene, der Syntax, ansetzt. Die Strukturebene, also der Einbezug der Semantik, bleibt hier offen. Damit ist dieses Modell hinsichtlich der Integrität eines Archivobjektes selbst zwar anzuwenden, kann jedoch so nicht für die Erhebung der Sicherheit für ein gesamtes Archiv ausreichend sein. Eine vollständige technische Sicherheitsanalyse ist so nicht machbar. Hier bedarf es einer dringenden Weiterentwicklung von Modellen, welche die Semantik mit einbeziehen.

2.2.8 Metadaten

Für eine vertrauenswürdige und abgesicherte Langzeitarchivierung digitaler Objekte müssen *Metadaten zu jeder Ebene* (physische, logische, konzeptuelle) generiert werden. Dies dient der vollständigen Erfassung von Information für ihre Erhaltung, ihrer Wiederauffindbarkeit und Verfügbarkeit. Metadaten stellen sicher, dass ein digitales Objekt im Archiv langfristig identifiziert und lokalisiert, angemessen verarbeitet und korrekt dargestellt werden kann. Metadaten sind entscheidend für die langfristige Verfügbarkeit und Benutzbarkeit der archivierten Information. Ohne Metadaten wäre eine digitale Langzeitarchivierung nicht möglich. Metadaten werden unterteilt in [Neu05]:

- Technische Metadaten
  - Technische Umgebung
  - Verarbeitungsprogramm
  - Dateiformattypen
- Strukturelle Metadaten
  - Relationen der digitalen Objekte in der Archivumgebung
- Beschreibende Metadaten
  - Handhabung der digitalen Archivobjekte (z.B. Identifikation, Lokalisierung, Kurzbeschreibungen)
- Administrative Metadaten
  - Lebenszyklus
  - Archivierungsmaßnahmen
- Rechte-Metadaten
  - Zugriffsmodalitäten
  - Rechtliche Verpflichtungen (Kopierschutz und Urheberschaft, Signaturgesetz)

2.2.9 Der digitale Bestand – Medien und Multimedia

Der digitale Bestand ist je nach Archiv unterschiedlich, besteht aber generell aus verschiedenen Medien, die wiederum in Medientypen, Formate und Formattypen unterschieden werden müssen, um deren Wechselbeziehungen darzustellen. Dies ist für die Gestaltung vertrauenswürdiger und abgesicherter Langzeitarchive von Bedeutung, da durch die bestandserhaltenden Maßnahmen Medienübergänge bzw. *Medienwechsel* auftreten, die auch als *Medienbrüche* mit Informationsveränderungen und -Verlusten behaftet sein können.

Der digitale Archivbestand einschließlich seiner unterschiedlichen Medien wird zunächst unterschieden in:

a. Selbst erstellte Objekte
b. Nach Vorgaben fremd erstellte Objekte

Selbsterstellte Objekte werden in direktem Bezug auf ihre künftige Verwendung nach entsprechenden Standards auf geeigneten Speichermedien mit präzisen Metadaten hergestellt und sind dementsprechend einsetzbar. Vorgaben sind von der jeweiligen Archivadministration individuell gegeben. Für die vertrauenswürdige und abgesicherte Langzeitarchivierung haben selbst erstellte Objekte den Vorteil, dass die Authentizität der Objekte einfacher sichergestellt werden kann als für fremd erstellte Objekte. *Fremderstellte Objekte* brauchen zudem übergreifende Austauschformate, wobei offene und freie Standards ebenso eine entscheidende Rolle spielen wie eine gemeinsame, möglichst standardisierte Metadatenstruktur.

2.2.10 Medien, Medientypen, Formate und Formattypen - Wechselbeziehungen

Medien, Medientypen, Formate und Formattypen sind Repräsentationsformen, Träger von Information und sind vom eigentlichen Inhalt der Information zunächst abgegrenzt zu definieren, was aber nicht ausschließt, dass sie diesen beeinflussen. Medien sind hier nicht die physischen Trägermedien sondern kennzeichnen eine bestimmte Art von Information wie in Tabelle 3 dargestellt. Das Format bzw. der

Formattyp entspricht der logischen Ebene eines digitalen Objektes. Das Medium bzw. der Medientyp spiegelt die konzeptuelle Ebene wieder. Die physische Ebene, wie von Thibodeau [Thi02] beschrieben, ist in dieser Unterteilung nicht enthalten.

Ein *Medientyp* ist als eine Klasse von bestimmten Medien definiert, der ein oder mehrere spezifische Medien angehören. Ein *Medium* gehört einer bestimmten Klasse Medientyp an, es ist ein Element einer, durch einen Medientypen definierte Klasse und ist in dieser spezifisch ausgezeichnet. Ein Medium wird durch sein konkretes *Format* spezifiziert. Dieses Format gehört wiederum einer, durch einen *Formattyp* definierten Klasse an. Je nach Art der Repräsentationsform und Kontext fungiert ein Format als konkrete Darstellungsform einer Information. In Tabelle 3 ist diese Differenzierung mit Beispielen dargestellt.

Tabelle 3: Zusammenhänge von Medientyp, Medium, Formattyp und Format.

Medientyp	Medium	Formattyp	Format
Alle Texte	Ein spezifischer Text	PDF, WORD, ASCII, XML, HTML, usw.	Konkrete Ausprägung eines Formattyps
Alle Bilder	Ein spezifisches Bild	BMP, JPEG, TIFF, SVG, usw.	Konkrete Ausprägung eines Formattyps
Alle Audio	Ein spezifisches Audiosample	WAV, MPEG, MIDI, usw.	Konkrete Ausprägung eines Formattyps

Auch wenn diese Definition keinen Bezug zu Inhalt und Kontext herstellt, ist ein solcher dennoch unweigerlich vorhanden. Denn je nach Bedeutung (Inhalt) und Anwendungsgebiet (Kontext) wird ein bestimmtes Format angewandt. So wird z.B. für den Medientyp Audio das Format in Abhängigkeit von dem damit zu repräsentierenden Inhalt wie Sprache, Rauschen, Musik (Gesang und verschiedene Instrumente), Instrumentalmusik, Geräusche oder Töne und der Umgebung ausgewählt. Das Gleiche gilt für alle anderen Medientypen. Änderungen des Formats, also Änderungen der Darstellungsform einer Information können sich auf ihren Inhalt auswirken und Ursache einer Informationsänderung sein. Gerde in Bezug auf Langzeitarchivierungssysteme ist dies ein Aspekt, der eine Herausforderung darstellt, denn geht durch Veränderungen des Formats Information verloren und sind folglich die Integrität und Authentizität der Information bedroht oder gar verletzt, ist die Vertrauenswürdigkeit der Langzeitarchivierungssysteme nicht mehr garantiert.

Format- bzw. Medienwechsel und Format- bzw. Medienbrüche

Als *Medienwechsel* wird die verlustfreie Transformation von Information von einem Medium auf ein anderes Medium bezeichnet. Ein Medium kann in verschiedenen Formaten vorliegen. Ein *Formatwechsel* ist die verlustfreie Transformation von Information innerhalb eines Mediums von einem Format auf ein anderes. Ein Medienwechsel impliziert gleichzeitig immer einen Formatwechsel.

Ein *Medienbruch* ist eine verlustbehaftete Überführung von Information eines Mediums auf ein anderes. Information geht dabei verloren. Liegt ein Medienbruch vor, ist dem immer die Annahme eines Medienwechsels vorangegangen. Dies gilt ebenso für die Definition von Formatwechsel und *Formatbruch*.

2.2.11 IT-Sicherheit

Medienbrüche können unterschiedlich verursacht werden. Aus Sicht der IT-Sicherheit ist bzw. sind dabei immer ein oder mehrere der folgenden Sicherheitsaspekte betroffen:

- **Verfügbarkeit** (*Availability*)
- **Vertraulichkeit** (*Confidentiality*)
- **Integrität** (*Data Integrity*)
- **Authentizität** (*Authenticity*)
- **Unleugbarkeit/ Nachweisbarkeit/ Nicht-Abstreitbarkeit** (*Non-Repudiation*)

Je nach Relevanz und Gewichtung des jeweiligen Sicherheitsaspekts ist ein Medienbruch mehr oder weniger schwerwiegend und mit einem Informationsverlust behaftet. Entsprechend mehr oder weniger leistungsfähig ist ein angewandter Sicherheitsmechanismus. An dieser Stelle soll die Aufführung der Sicherheitsaspekte nur eine kurze Einführung geben. Auf die Rolle der IT-Sicherheit für vertrauenswürdige und abgesicherte Langzeitarchive wird ausführlich in Kapitel 5 eingegangen.

2.2.12 Erhaltungsmaßnahmen

Zu den grundsätzlichen Erhaltungsmaßnahmen in einem Langzeitarchivierungssystem gehören:

- An die Technik angepasste und angemessene Speicherung
- Identifizieren und Wiederauffinden aller Komponenten eines digitalen Objektes (Inhalt, Speicherung, Darstellung) einschließlich der dazugehörigen zusätzlichen Metadaten
- Korrekte Verarbeitung der digitalen Objekte, so dass keine Information verloren geht
  - Interpretation des Bitstroms
  - Komposition von digitalen Objekten auf seinen verschiedenen Ebenen
  - Richtige Darstellung

Bedingungen

Aus den aufgeführten Erhaltungsmaßnahmen können die folgenden *Bedingungen für die abgesicherte Langzeitarchivierung digitaler Objekte* abgeleitet werden:

Die Erhaltung und Bereitstellung/ Zugriff sollten nicht getrennt gehandhabt, sondern miteinander verbunden werden. Durch solch einen Mechanismus kann die korrekte Erhaltung eines digitalen Objekts im gleichen Atemzug mit seiner Auslieferung überprüft werden. Veränderungen, die durch Bestandserhaltungsmaßnahmen wie etwa der Migration verursacht werden, müssen die Erhaltung von digitalen Objekten im Langzeitarchiv nicht notwendigerweise negativ beeinflussen. Aufgrund der Schnelllebigkeit der Medien, muss die Speicherung der digitalen Objekte den aktuell geltenden Medien angepasst werden. Dies erfordert Maßnahmen wie z.B. Transformation, damit das Objekt nicht unzugänglich wird. Zur Bestandserhaltung werden digitale Archivobjekte dazu auf physischer und logischer Ebene verändert, was jedoch in der Regel die konzeptuelle Ebene nicht beeinflussen sollte. Die Schwierigkeit besteht darin, angemessene Änderungen eines digitalen Objektes für seine Erhaltung zu bestimmen. Je nach Langzeitarchiv und seinem Einsatzgebiet muss zwischen den folgenden Fragen abgewogen werden:

- Welche Veränderungen am digitalen Objekt sind ohne gravierende Auswirkungen?
- Welche Veränderungen am digitalen Objekt haben unvertretbare Auswirkungen und sind daher ausgeschlossen?
- Welche Veränderungen sind notwendig und am meisten von Vorteil?

Auf diese Fragen können keine allgemeingültigen Antworten gegeben werden. Grundsätzlich gilt aber, dass das *Ziel*, der *Zweck*, die *Umgebung* und der *Gegenstand* eines Langzeitarchivs betrachtet werden müssen, um die Bedingungen und daraus abgeleitet die *Anforderungen* an das Langzeitarchiv zu bestimmen. Die vollständige Erhaltung der Authentizität, was bedeutet, dass der Output unverändert identisch ist mit dem Input, ist in digitalen Langzeitarchiven nicht oder nur schwer zu realisieren, da die Technik einem ständigen Wandel unterliegt.

Die Wahl der *geeigneten Strategien* zur digitalen Bestandeserhaltung ist abhängig von dem Ziel, dem Zweck, der Umgebung und des Gegenstand eines Langzeitarchivs einschließlich der Art der zu archivierenden Information. Vertrauenswürdige und abgesicherte Langzeitarchive schließen dabei die Sicherung der Integrität und Authentizität der Information ein. Dies bedeutet, dass das Sammelgut je nach Vorgaben in der Substanz/ Inhalt unverändert erhalten bleiben soll, dass ein Schutz vor illegalen Manipulationen gewährleistet werden soll und dass die Urheberschaft und Unversehrtheit gesichert werden soll.

2.2.13 Zusammenfassung resultierender Anforderungen an vertrauenswürdige und abgesicherte digitale Langzeitarchive

Den vorhergehenden Ausführungen zur Folge können Anforderungen an vertrauenswürdige und abgesicherte Langzeitarchive digitaler Informationen wie folgt zusammengefasst werden:

Generelle Anforderungen an die Langzeitarchivierung

- Langfristige Sicherheit
- Langfristige Überprüfbarkeit der Integrität, Authentizität
- Langfristige Lesbarkeit
  - des digitalen Objekts
  - der Sicherheitsmechanismen
  - gesicherte (signierte) Dokumente

Qualitätskriterien

Die folgenden aufgeführten Anforderungen dienen als Qualitätskriterien zur Festlegung der generellen Ziele zukünftiger Systeme zur vertrauenswürdigen Langzeitarchivierung:

- Interoperabilität
- Skalierbarkeit
- Modularisierung der technischen Infrastruktur
- Flexibilität und Unabhängigkeit in Bezug auf Hard- und Softwaretechnologien
- Vertraulichkeit
- Offenheit und Transparenz
- Benutzbarkeit
- Verfügbarkeit
- Revisionssicherheit
- Vertrauenswürdigkeit

Allgemeine Anforderungen

Daraus abgeleitete allgemeine Anforderungen für Systeme der Langzeitarchivierung digitaler Objekte sind:

- **Realisierbarkeit:**
  - Hardware und Software müssen in der Lage sein, die Verfahren umzusetzen.
  - realisierbare und kostenvernünftige Umsetzung.
- **Nachhaltigkeit:**
  - Intern: Systeme und Verfahren müssen gegenüber Auswirkungen von technologischen Veralterungen immun sein und davon isoliert werden können.
  - Extern: Systeme und Verfahren müssen miteinander verbunden und bzw. verkettet aufeinander abgestimmt werden können.
- **Anwendbarkeit:**
  - Die Umsetzung eines Langzeitarchivierungssystems muss sich innerhalb von angemessenen und sinnvollen Grenzen bzgl. der Schwierigkeit/ Komplexität und der Kosten bewegen.
- **Angemessenheit:**
  - Ist abhängig von der Art der digitalen Objekte und der objektiven Sachlage einschließlich der Anforderungen in Bezug auf ihre Erhaltung.

  - Bandbreite von Möglichkeiten, aus denen eine angemessene Lösung gewählt werden kann.
- **Veränderbarkeit:**
  - Es müssen vertretbare Veränderungen am digitalen Objekt festgelegt werden.

Anforderungen an die Handhabung digitaler Objekte

Zur Handhabung digitaler Objekte im Archiv gilt ein Grundsatz an Anforderungen, der im Folgenden aufgelistet ist:

- Angemessene Sicherung der
  - Nachweisbarkeit
  - Integrität
    - Physische Ebene
    - Logische Ebene
    - Konzeptuelle Ebene
  - Authentizität
  - Verfügbarkeit
  - Vertraulichkeit
- Einhaltung von Normen und verfügbaren Standards
- Sicherstellung der Lesbarkeit zu jedem angeforderten Zeitpunkt
- Benutzbarkeit
- Vollständigkeit
- Nachvollziehbare erkennbare Veränderungen (Vervollständigungen, Änderungen) und Urheberbestimmung
- Management des Lebenszyklus (*Preservation Planning*)
- Medienwechsel ohne Informationsverlust

Funktionale Anforderungen

Funktionale Anforderungen beziehen sich auf die Handhabung digitaler Objekte, stehen aber gleichzeitig im Zusammenhang mit den Funktionsentitäten, wie sie im OAIS-Referenzmodell festgelegt sind. Sie können wie folgt zusammengefasst werden:

- Geregelter und autorisierter Zugriff auf einzelne oder mehrere digitale Objekte für eine bestimmte Benutzergruppe
- Datenbankgestützte Verwaltung der digitalen Objekte mit Hilfe von Metadaten, ggf. Volltexterschließung der Inhalte der archivierten Objekte
- Unterstützung verschiedener Indizierungs- und Recherchestrategien zur Gewährleistung des direkten Zugriffs auf Information
- Einheitliche und gemeinsame Speicherung beliebiger Information als digitale Objekte
- Anpassbarkeit der Datenbanksysteme an Änderungen der Gesetzeslage
- Verwaltung und Zugriffsregelung von Speichersystemen
- Sicherstellung der Verfügbarkeit der gespeicherten Information über einen längeren Zeitraum
- Bereitstellung von digitalen Objekten unabhängig von der sie ursprünglich erzeugenden Anwendung auf verschiedenen Systemen
- Vereinfachung der Erfassung von Information als digitale Objekte durch Unterstützung von Klassen-Konzepten; Vererbung von Merkmalen und Strukturierung der Informationsbasis
- Erzeugung von langfristig stabilen und einheitlichen Archivformaten und Abspielsystemen zur Darstellung von digitalen Archivobjekten, unabhängig von der sie ursprünglich erzeugenden Anwendung (Konverter)

- Absicherung der digitalen Archivobjekte im Archiv gegen unberechtigten Zugriff und gegen Veränderbarkeit zu jedem Zeitpunkt im Archiv
- Übergreifende Verwaltung unterschiedlicher Speichersysteme und schneller Zugriff und Bereitstellung der Information durch Zwischenspeicher (*Caches*)
- Standardisierungen
  - Schnittstellen zur Integration digitaler Archive als Dienste in beliebige Anwendungen
  - Formate für eine langfristige Verfügbarkeit und Sicherheit
  - Metadaten für eine langfristige Verfügbarkeit und für Migrationssicherheit
- Automatisierte und eigenständige Wiederherstellungsfunktionalität (*Recovery*); verlustfreie Rekonstruktion inkonsistenter oder gestörter Systeme
- Nachvollziehbarkeit von Handlungen/ Aktivitäten
  - sichere Protokollierung von Veränderungen an Strukturen und Digitalen Objekten/ Informationspaketen sowie deren Verarbeitung
  - Sicherstellung von Konsistenz und Wiederauffindbarkeit
- Automatisierte, nachvollziehbare und verlustfreie Verfahren (Migration/ Emulation) zum Preservation Planning

Rechtliche Anforderungen

Die rechtlichen Anforderungen müssen ebenso in einem System zur Langzeitarchivierung digitaler Information integriert sein. Sie spielen eine zentrale Rolle, da sie Sicherheitsbedrohungen benennen und teilweise verschärfen. So müssen beispielsweise bei der Auslieferung die Urheberrechte genaue beachtet werden, was bedeutet, dass dafür wirksame Mechanismen vorhanden sein müssen, die solche rechtlichen Aspekte im Langzeitarchivierungssystem integrieren. Rechtliche Aspekte können wie folgt aufgelistet werden:

- Aufbewahrung
  - Spätere Verfügbarkeit
  - Andauernde Rechtssicherheit
  - Kontrollzwecke und Rechenschaft
  - Langfristige bzw. dauerhafte Erhaltung der Inhalte von allgemein öffentlichem Interesse und als Quellengrundlage für vielfältige wissenschaftliche Forschungen
- Nachvollziehbarkeit/ Nachweis bestimmten Handelns oder Unterlassens sowie eines Handlungsverlaufs
- Fälschungssicherheit
- Vollständigkeit
- Ggf. Einsatz als Beweismittel
- Lesbarkeit unabhängig vom Erstellungsmedium
- Einhaltung der Urheberrechtsbestimmungen
- Einhaltung der Gesetzeslage und aktuellen gesetzlichen Vorschriften
- Anpassung an Änderungen der Gesetzeslage

2.2.14 Annahmen

Der vorliegenden Expertise liegt das OAIS-Referenzmodell als formales Modell und die Definition des digitalen Objektes laut Thibodeau als formale Beschreibung von digitalen Objekten zugrunde. Dadurch kann ein allgemein geltendes und standardisiertes Vergleichsmaß für die Systemabstraktion und Evaluation der Beispielsysteme erlangt werden. Darüber hinaus bietet dies die Grundlage für die Ermessung des Handlungsbedarfes.

Die Betrachtung von *Safety* wird in dieser Expertise ausgeschlossen und es wird sich auf die Betrachtung von *Security* für die digitale Langzeitarchivierung beschränkt.

In dieser Studie wird sich auf die Sicherheitsanalyse

a) der Systemkomponenten, ausgehend von der Erhebung der Systemarchitektur, sowie

b) des digitalen Archivobjektes, ausgehend von der Erhebung der Informationsflüsse

bezogen. Die Sicherheitsanalyse wird demnach zum einen systemorientiert und zum anderen objektorientiert erfolgt. Der dortige Handlungsbedarf wird aufgezeigt. Eine szenarienabhängige und systemspezifische Analyse würde eine Netzwerksicherheitsanalyse mit expliziten und detaillierten Security Scans erfordern, was in dieser Studie nicht vorgesehen ist. Die Systemarchitektur und der Informationsflüsse werden abstrahiert erhoben, um anwendbare Sicherheitsmechanismen über die Anwendungsszenarien hinweg bestimmen zu können. Dadurch lassen sich allgemeingültige Aussagen treffen über einen generellen Handlungs- und Standardisierungsbedarf, denn Ziel dieser Studie ist es, den bestehenden nestor-Kriterienkatalog, der allgemeingültig ist für alle Arten von Langzeitarchiven, im Punkt Sicherheit zu erweitern und Vorschläge für zukünftige Weiterentwicklungen zu präsentieren.

3 Analyse der Beispielszenarien und allgemeine Charakterisierung der exemplarischen Langzeitarchivierungssysteme

In diesem Kapitel werden *exemplarische Langzeitarchivierungssysteme*, wie sie in zwei verschiedenen Anwendungsszenarien zum Einsatz kommen, allgemein charakterisiert. Digitale Langzeitarchivierung ist ohne solch eine Kontextmodellierung mit Bezug auf eine spezielle Nutzergruppe und Anwendungs-Domain nicht möglich. Nur so können semantische Verknüpfungen erfasst und eine Struktur erkannt werden, um eine Absicherung des Archivs im Ganzen zu ermöglichen, also sowohl funktional das Archivobjekt betreffend als auch auf struktureller Archivebene. Auf diese Weise können organisatorische Aspekte mit einbezogen werden. Aber auch zeitliche gesellschaftsbedingte Gegebenheiten und technische Entwicklungsstandards können so abgebildet und erfasst werden.

Die Absicherung eines digitalen Langzeitarchivs kann also nur im *Kontext* dessen Anwendung und der Nutzergruppe geschehen, da je nach Einsatzgebiet und Community verschiedene Prozesse, Komponenten, Ressourcen sowie unterschiedliche Informationsflüsse vorherrschen. Folglich bestehen auch unterschiedliche Anforderungen an und Annahmen über an ein Langzeitarchivierungssystem. Aus diesem Grund werden hier im Folgenden Akteure, Rollen, Systemkomponenten und Ressourcen sowie Informationsflüsse in zwei beispielhaften Szenarien beschrieben. Dies soll die semantischen Dimensionen aufzeigen, die bei der Entwicklung vertrauenswürdiger und abgesicherter Langzeitarchivierungssysteme digitaler Multimediainhalte zu berücksichtigen sind. In dieser Expertise wurden Anwendungsszenarien gewählt, welche zwei der wichtigsten, repräsentativsten und erfahrensten Nutzergruppen von Langzeitarchivierungssystemen digitaler Multimediainformationen repräsentieren: **Öffentlich-rechtliche Rundfunkanstalten** und **Hochschul-Medienzentren**. Beide Szenarien haben die Aufgabe, Kulturgüter und Information von öffentlichem Interesse zu erhalten und bereitzustellen. Die Auswahl dieser Szenarien ermöglicht es, auf das vorhandene Know-how zurückzugreifen, um darauf aufbauend für die Zukunft allgemeine Empfehlungen für die Absicherung digitaler Langzeitarchivsysteme multimedialen Inhalts erstellen zu können.

Die Szenarien dienen somit als *praktische Beispiele*, an denen die Integration von IT-Sicherheit illustriert wird. Dazu werden die grundsätzlichen Komponenten der allgemeinen technischen Infrastruktur in Funktion und Struktur sowie die Informationsflüsse erhoben, um darauf aufbauend die Gefahren darzustellen und im folgenden Kapitel die Integrationsmöglichkeiten von IT-Sicherheit zu evaluieren, die für vertrauenswürdige und abgesicherte Langzeitarchivierungssysteme eingesetzt werden können und müssen.

3.1 Öffentlich-rechtliche Rundfunkanstalten

In Deutschland besteht das *Duale Rundfunksystem*, in welchem öffentlich-rechtliche Anstalten und private Rundfunkanbieter parallel die Öffentlichkeit mit Informationen versorgen.

Während sich das Handeln des privaten Rundfunks vornehmlich an der Maximierung der Einschaltquote innerhalb einer selbst gewählten Zielgruppe orientiert hat der öffentlich-rechtliche Rundfunk die *Grundversorgung* sicherzustellen. Die öffentlich-rechtlichen Sender sind an den Programmauftrag aus

dem *Rundfunkstaatsvertrag (RStV)* gebunden: „Der öffentlich-rechtliche Rundfunk hat durch die Herstellung und Verbreitung von Hörfunk- und Fernsehprogrammen als Medium und Faktor des Prozesses freier individueller und öffentlicher Meinungsbildung zu wirken." [RStV06, §11(1)] Bei der Erfüllung dieses Auftrages sind die Grundsätze der Objektivität und Unparteilichkeit, der Meinungsvielfalt sowie der Ausgewogenheit zu berücksichtigen [RStV06, §11(3)]. Das Programm hat der Information, Bildung, Beratung und Unterhaltung zu dienen [RStV06, §11(2)].

Dem öffentlich-rechtlichen Rundfunk ist gestattet, programmbegleitende Druckwerke und Telemedien mit programmbezogenem Inhalt anzubieten. Die öffentlich-rechtlichen Rundfunkanstalten unterhalten im Internet eine Reihe von Internetangeboten.

Zu den öffentlich-rechtlichen Anstalten zählen:

- Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland (ARD) als Zusammenschluss der Landesrundfunkanstalten
  - Bayerischer Rundfunk (BR), München
  - Hessischer Rundfunk (hr), Frankfurt am Main
  - Mitteldeutscher Rundfunk (MDR), Leipzig
  - Norddeutscher Rundfunk (NDR), Hamburg
  - Radio Bremen (RB), Bremen
  - Rundfunk Berlin-Brandenburg (RBB), Berlin und Potsdam
  - Saarländischer Rundfunk (SR), Saarbrücken
  - Südwestrundfunk (SWR), Stuttgart
  - Westdeutscher Rundfunk (WDR), Köln
- Zweites Deutsches Fernsehen (ZDF)
- Deutsche Welle
- DeutschlandRadio (DLR)

3.1.1 Rundfunkarchive

Die Deutsche Gesetzgebung verlangt, dass gesendetes Material grundsätzlich für mindestens 100 Tage aufzubewahren ist [Chr06, S. 609]. Die Aufzeichnungen dürfen nicht veränderbar sein.

Die öffentlich-rechtlichen Rundfunkanstalten haben sich „verpflichtet, für die Überlieferung ihres Programmvermögens selbst zu sorgen." [BR04] Sie betreiben eigenständige Archive.

Die *Ziele der Archive* der Rundfunkanstalten stellen sich wie folgt dar [BR04]:

- **Archivierung und Dokumentation** In den Fernseharchiven „werden Produktionsmaterialien und gesendete Produktionen" der von den Rundfunkanstalten „veranstalteten bzw. belieferten Programme archiviert und dokumentiert. Die Bestände" der Fernseharchive „bilden einen einzigartigen kultur- und zeitgeschichtlichen Fundus und sind Teil des kulturellen Erbes."
- **Nutzbarmachung für Wiederverwertung** „Die Herstellungskosten für eine Minute Fernsehen liegen so hoch, dass Archivmaterialien höchsten ökonomischen Wert für Programmveranstalter besitzen. Archivbestände bilden daher ein unersetzliches Programmvermögen und einen wichtigen Fundus für Programmschaffende."

Die Archive der öffentlich-rechtlichen Rundfunkanstalten sind Archive im traditionellen Sinne. Klar abzugrenzen ist die Verwendung des Begriffes in der Informationstechnik, welche Archive auf die systematische *Erfassung* und *Verwahrung* reduziert. Rundfunk-Medienarchive dienen darüber hinaus der *Ordnung, Verwaltung und Verwertung* der Archivalien. Bildlich formuliert sind sie ein Ort, „wo Daten leben" und nicht „wo Daten sterben" [Wri07, S. 132].

Aufgrund ihres Doppelauftrages – Archivierung und Dokumentation sowie Nutzbarmachung für Wiederverwertung – sind die Archive in den Produktions- und Sendeprozess eingebunden und treten als vielfältiger *Dienstleister* auf [BR04]:

- Verwaltung von Dreh- und Arbeitsmaterialien aus Produktion und Sendung. Formalerfassung gesendeten Programmmaterials.
- Sachgerechte Lagerung, Verwaltung und Ausleihe des Materials. Verantwortlicher langfristiger Erhalt des Programmvermögens.
- Erstellung eingehender Beschreibungen der Sach- und Bildinhalte des Materials. Auftragsrecherche verlangter Bildfolgen, Inhalte und Beiträge.
- Wechselseitiger Austausch von Material mit anderen öffentlich-rechtlichen Rundfunkanstalten.
- Bereitstellung von Bildern und Tondokumenten aus dem Archiv. Verwaltung temporär archivierten Fremdmaterials.

„Rundfunk-Medienarchive werden hauptsächlich dafür eingesetzt, um deren Inhalte wieder zu verwenden." [Wri07] So trägt das Archiv der BBC zu ungefähr 30% zu den Fernsehnachrichten bei und im Laufe eines Jahres werden etwa insgesamt 20% des Archivbestandes angefordert. Diese Quote lässt sich voraussichtlich steigern bei einer konsequenten Weiterentwicklung zu einem direkten, öffentlichen Zugriff auf die Inhalte.

Durch den hohen Stellenwert der Wiederverwendung von Programmmaterial treten Archive nicht ausschließlich als Endlager am Ende der Produktionskette in Erscheinung. Die Archive der Rundfunkanstalten sind eng in den Produktionsprozess von Fernseh-Programmbeiträgen eingebunden.

3.1.2 Produktion von Fernsehbeiträgen

Fernsehsendungen unterscheiden sich in Unterhaltungssendungen, Nachrichtensendungen und Bildungsfernsehen [KK05, S. 103]. Als *Formate* können Dokumentationen und Reportagen, Nachrichten, Magazine, Fernsehfilm, Shows, Sportsendungen und weitere unterschieden werden.

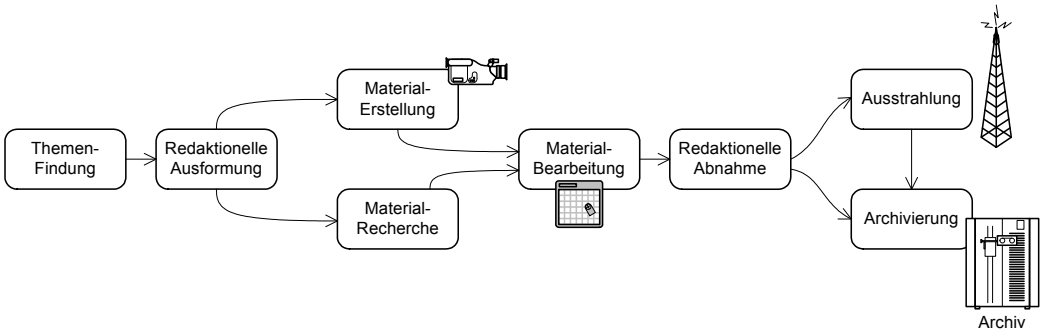

Abbildung 7: Allgemeiner Produktionsprozess für eine aktuelle Berichterstattung, vgl. [San05, S. 110], [Ebn05, S. 6], [Ott05].

Der Produktionsprozess für eine aktuelle Berichterstattung stellt sich beispielsweise wie folgt dar (Abbildung 7), vgl. [San05, S. 110], [Ebn05, S. 6], [Ott05]: Zu Beginn steht die Themenfindung. Die Redaktion trifft hierzu eine Auswahl aus Ereignissen und Ideen, welche sie Agenturmeldungen, Zuschauerreaktionen und weiteren Quellen entnimmt. In der sich anschließenden Ausformung ergänzen Redakteure mittels Hintergrundrecherchen die Inhalte und planen die Gestaltung und weitere Herstellung des Beitrages. Sofern möglich werden Archivbestände in der aktuellen Produktion wieder verwendet. Ein Kamerateam dreht vor Ort neues Bild- und Tonmaterial. Alternativ kommt von der meldenden Agentur zugespieltes Material zum Einsatz. Für sämtliches Material ist sicherzustellen, dass die Nutzungsrechte vorliegen. Ergänzend erstellt die Produktion Ton- und Sprachspuren sowie Grafikelemente. Die Nachbearbeitung fügt im Videoschnitt dieses Rohmaterial zum Beitrag zusammen. Nach dessen Prüfung und Abnahme durch die Redaktion wird der Fernsehbeitrag in den Sendeplan eingegliedert und kommt zur Ausstrahlung, um abschließend im Archiv katalogisiert dauerhaft aufbewahrt zu werden.

Der gesamte Produktionsprozess wird bestimmt von dem Image des Fernsehsenders, dem Format der Sendung sowie der Programmplanung, in welche der Beitrag eingegliedert wird.

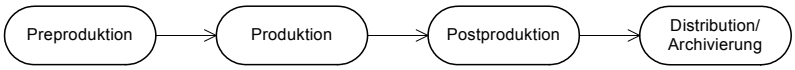

Abbildung 8: Medienproduktionsprozess, vgl. [KK05, S. 19f].

Allgemein unterteilt sich *Medienproduktion* in Preproduktion, Produktion, Postproduktion und Distribution/ Archivierung (Abbildung 8) [KK05]. Diese Einteilung ist ebenfalls auf die Produktion von Fernsehbeiträgen anwendbar.

1. **Preproduktion** Planung und Recherche von Inhalten
2. **Produktion** Erstellung der Inhalte
3. **Postproduktion** Verfeinerung, Bearbeitung, Test
4. **Distribution/ Archivierung** Übergabe der Inhalte an die Zielgruppe und Aufbewahrung

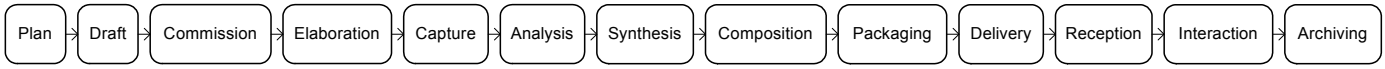

Abbildung 9: Linearer Arbeitsablauf in der traditionellen Fernsehproduktion (Linear Push Workflow Model), vgl. [MT04].

„Bei der Analyse von Prozessen verschiedener Rundfunkanstalten hat sich gezeigt, dass sehr viele Gemeinsamkeiten zu finden sind. Sie werden zur Definition von allgemein gültigen generischen Prozessen genutzt." [EK05, S. 15]

Traditionell beginnt der *Herstellungsprozess* (Abbildung 9) mit einem Planungs- und Entwurfsabschnitt (*Plan, Draft*), in welchem die Ursprungsidee herausgebildet und als Projekt skizziert wird. Im Falle eines Umsetzungsbeschlusses (*Commission*) folgt die Ausarbeitung der Projektdetails (*Elaboration*) mitsamt der Planung der tatsächlichen Herstellung. Der Abschnitt der Produktion leistet die Erstellung der Inhalte durch Dreharbeiten für Neuaufnahmen und die Erschließung von Bestandsmaterial (*Capture*) bei anschließender Sichtung und Auswertung (*Analysis*). Die Postproduktion (*Synthesis, Composition, Packaging*) überarbeitet das Rohmaterial und fügt dieses zu dem fertigen, sendefähigen Beitrag zusammen. Über eine Ausstrahlung oder andere Verbreitungswege gelangt der Beitrag zu den Zuschauern (*Delivery*). Mit der Archivierung (*Archiving*) endet die Produktionskette. [MT04]

Der allgemeine Produktionsprozess zur Erstellung eines Fernsehbeitrages ist unabhängig von der eingesetzten Technik und angewendeten Technologie [San05].

3.1.3 Wandel der Rundfunkanstalten zur digitalen, verteilten und vernetzten Fernsehproduktion

„Um die inzwischen erheblich gestiegenen Anforderungen an die Fernsehproduktion auch in Zukunft in vollem Umfang bewältigen zu können, sind die Rundfunkanstalten zu einer umfassenden Neustrukturierung ihrer Produktionstechnik und Arbeitsabläufe gezwungen. Die angestrebte Effizienzsteigerung ist nur durch weitere Digitalisierung und Automatisierung der Produktionsmittel und -abläufe möglich. Das bedeutet, dass bei der Programmproduktion im Studio, bei der Nachbearbeitung und bei der Archivierung zukünftig verstärkt Computer- und Netzwerktechnologien eingesetzt werden."[3]

Die öffentlichen Rundfunkanstalten stellen sich dieser Herausforderung und befinden sich derzeit in einem umfangreichen *Umstellungsprozess*. Die Technik hat sich in Richtung Digitalisierung entwickelt [Sau07]. „Insgesamt ist zu beobachten, dass die klassische Speicherung auf Videoband zunehmend, umfassend und unaufhaltsam durch Netzwerke mit Servern ersetzt wird." [EK05, S. 14]

Zwei sich ergänzende Entwicklungen bestimmen diesen technischen Wandel:

[3] IRT, Sachgebiet Produktionssysteme Fernsehen, http://www.irt.de/de/irt/organigramm/produktionssysteme-fernsehen.html, 21.08.2007

- Digitale Aufzeichnung der Video- und Audioinhalte
- Einsatz von Computer- und Netzwerktechnik in der Programmproduktion

Der Wandel in den Rundfunkanstalten umfasst ebenfalls eine Umorganisation der Arbeitsabläufe. Sie werden an die auf Informationstechnik gestützte Produktionsweise angepasst [Sau07]. Mit dem technischen Wandel verbunden ist ebenfalls eine Umstellung des Arbeitsmodells.

Der traditionelle Produktionsablauf ist durch eine lineare Aufeinanderfolge der Schritte gekennzeichnet (*Linear Push Workflow Model*) (Abbildung 9) [MT04], [Sau07]. Das Arbeitsmaterial durchläuft in Abfolge die einzelnen Schritte des Herstellungsprozesses. Es steht wegen seines physischen Trägermaterials lediglich einmalig zur Verfügung, eine parallele Mehrfachnutzung erfordert zusätzliche Kopien der Inhalte. Metadaten werden manuell geführt. Der Transport der Inhalte zwischen den Arbeitsschritten erfolgt mechanisch auf Datenträgern. Lediglich Inseln der Produktion sind digitalisiert. Weil die Systeme nicht dafür ausgelegt sind, Metadaten untereinander gemeinsam zu nutzen geht ein nicht unerheblicher Teil der Information über den Herstellungsprozess verloren und muss zur Archivierung händisch neu ermittelt und in den Katalog eingepflegt werden.

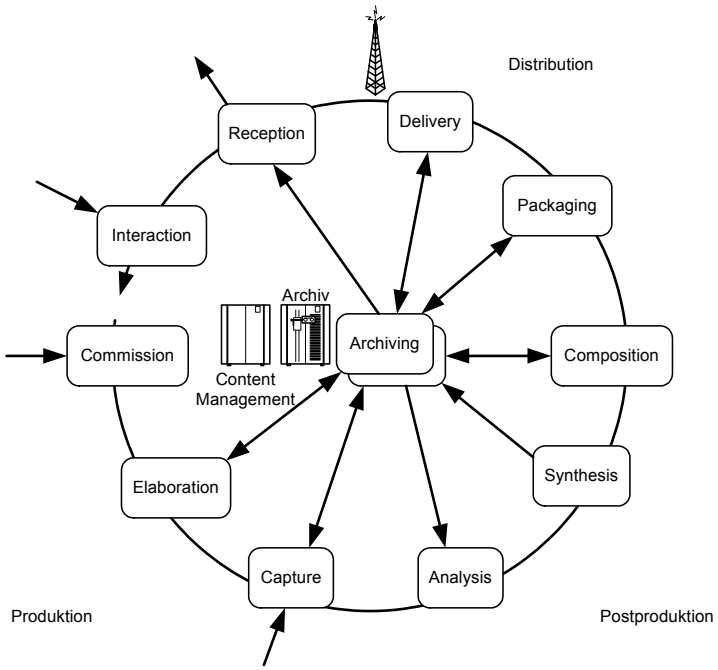

Abbildung 10: Am Inhalt ausgerichteter Arbeitsablauf in der Fernsehproduktion (Concent Centric Pull Workflow Model), vgl. [MT04], [Sau07, S. 297].

In der neuen Organisation (Abbildung 10) richtet sich die Produktionsfolge strikt am Inhalt aus (Content Centric Pull Workflow Model) [Sau07]. Sie folgt der *Philosophie der ständigen Informationsanreicherung entlang der Produktionskette*. Eine zentrale Instanz ist für die Verwaltung der Inhalte einschließlich Metadaten zuständig und dient als Verteiler. Alle angeschlossenen Produktionsstufen können Einsicht nehmen in den Bearbeitungsfortschritt von Material und dieses anfordern. Nach einer Bearbeitung werden Änderungen unmittelbar einpflegt. Während die zentrale Verwaltung nach Möglichkeit sämtliche im Herstellungsprozess anfallenden Informationen vorhält erhalten die Produktionsstufen den für sie jeweils relevanten Ausschnitt.

Einigkeit besteht darin, dass ein Content-Management die Inhalte verwaltet und bereitstellt. Einige Quellen setzen diese zentrale Aufgabe mit den Archiven [Sau07] gleich. Ihnen kommt eine bestimmende Rolle zu als „zentrale, integrale Bibliothek der Produktion." Sie sind nicht länger das „Endlager der Sendung".

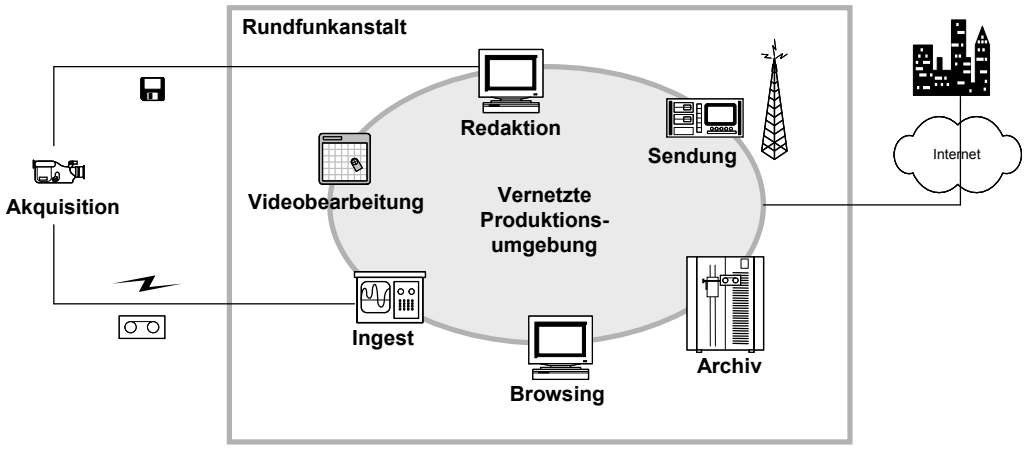

Abbildung 11: Grundsätzliche Gliederung der allgemeinen technischen Infrastruktur der Rundfunkanstalten bei Datei-basierte Fernsehproduktion, vgl. [EK05, S. 16].

Die zukünftige Organisation sieht eine *Vernetzung der Produktionsabschnitte* vor (Abbildung 11). Der Großteil der Fernsehproduktionen erfolgt von Anfang bis Ende durchgängig gestützt auf Informationstechnik. Inhalte stehen unmittelbar nach ihrer Anlieferung allen Redaktionen zur Verfügung und können parallel genutzt werden. Die Übertragung des Materials erfolgt über reguläre Informationstechnik-Datennetze.

Das angestrebte Ziel der Archive innerhalb der Rundfunkanstalten lautet, das vorhandene Material vollständig zu erschließen. „Die vorrangigste Bemühung der Rundfunkarchive ist es, einen einzigen digitalen Speicherort aufzubauen, der dann alle Geschäftsbereiche bedient.“ [Wri07] Die geringen Kosten für Speichersysteme erlaubten es, lokale Ablagen der Inhalte – zumeist in proprietären Formaten der Einzelkomponenten – anzufertigen. Derartige *lokale Privatarchive* sind jedoch „der Feind der Archive“, denn sie bewahren das Material lediglich für eine Person oder eine überschaubaren Gruppe. Rundfunkveranstalter profitieren hingegen von der Wiederverwendung ihres Materials. Zu diesem Zweck wurden Rundfunkarchive geschaffen.

„In der Übergangszeit von der Speicherung auf Kassetten zur Datei-basierten Technik werden Mischformen unumgänglich sein. Dies wiederum bedingt eine Vielzahl von Medienbrüchen.“ Die vorübergehende Koexistenz herkömmlicher und IT-gestützter Techniken schafft Übergangsprobleme [Sau07].

Heterogenität

Indizien weisen darauf hin, dass die öffentlichen Rundfunkanstalten einen hohen Grad an Autonomie besitzen.[4] In Folge haben sich in den Rundfunkanstalten *Einzellösungen* für die Organisation herausgebildet mit einer Vielfalt an Formaten, technologischer Infrastruktur und Prozessen in Produktion und Archivierung.

„Die Standardsoftware für die Rundfunktechnik gibt es schlechthin nicht.“ [Sau07, S. 298] Die Lösungen innerhalb der untersuchten öffentlich-rechtlichen Rundfunkanstalten bestätigen diese Feststellung. Trotz der Zielsetzung einer gleichartigen Technik aus einer Anbieterhand entschlossen sich Planer unter anderem für einen Verbund integrierter Teilsysteme, welche für ihren jeweiligen Einsatzbereich die Besten ihrer Art darstellen. Evaluationen hatten offenbart, dass keines der marktüblichen Komplettsysteme die geforderten Funktionalitäten in allen Bereichen hinreichend abdeckte. Die Integration der Teilsysteme bedingt Anpassungen der Teilsysteme sowie eine Neuentwicklung zahlreicher Schnittstellenformate und Vermittlungsmechanismen zwischen den Teilsystemen. Dass in der vernetzten Produktionsumgebung die Teilsysteme vielfältig zueinander in Beziehung stehen erhöht die Anzahl der für diese Konstellation spezifischen Abstimmungsmaßnahmen zusätzlich.

[4] Fehlende einheitliche Regelung für Metadaten [Ebn05, Inhaltsangabe]; Variierende technische Infrastruktur für die Verwaltung von Rechteinformation [NN01]; Abbildungsverfahren für die Kopplung von Rechteinformation [NN01]; Individuelle Datenbanken der Rundfunkanstalten zu Rechteinformation [NN01, S. 82]; Festgestellte Gemeinsamkeiten in den analysierten Prozessen verschiedener Rundfunkanstalten [EK05, S. 15].

Der gegenwärtige Wandel in den Sendeanstalten bietet Gelegenheit für eine Vereinheitlichung der verwendeten Infrastruktur. „Für die reale Systemintegration von IT-gestützten Produktionskomponenten ist es sinnvoll, gerade im Hinblick auf eine Minimierung des Aufwandes und somit der Kosten, einheitliche Formate zu betrachten und einzuführen." [EK05, S. 15] Ob diese Gelegenheit ergriffen wird ist noch offen. Leider ist festzustellen, dass umfangreiche Abschnitte des Produktionsprozesses auf einem einzigen System zusammengefasst und von einem Hersteller angeboten werden [Sch05]. Jedoch ist inzwischen eine wachsende Bereitschaft der Hersteller an vereinheitlichten Lösungen zur Systemintegration zu beobachten [EK05, S. 17]. Eine modulare, offene Systemarchitektur mit definierten Schnittstellen ist eine unabdingbare Forderung für die auf Informationstechnik gestützte Produktionsumgebung. Offene Standards schaffen Investitionssicherheit und Herstellerunabhängigkeit, Zuverlässigkeit und Flexibilität. Die Hersteller bieten aber ausschließlich einzelne Großsysteme an [Sau07]. Einzellösungen werden auch weiterhin teuer und wenig effektiv sein. Herstellerabhängigkeiten sind zu vermeiden. „Die *Interoperabilität* wird zu einem entscheidenden Faktor in der IT-gestützten Programmproduktion." [Sau07, S. 298]

3.1.4 Akteure in der digitalen, vernetzten Produktion

Als bedeutende *Akteure* in dem Erstellungsprozess für audiovisuelle Inhalte können ausgemacht werden:

- Programm-/ Produktionsplanung
- Zuspielung/ Akquisition
- Logging
- Redaktion
- Dokumentation
- Honorare-&-Lizenzen
- Post-Produktion
- Sendung

3.1.5 Rundfunkarchive in der digitalen, vernetzten Fernsehproduktion

Digitale Archive müssen folgenden Mindestsatz an *Operationen* anbieten [Wri07]:

- **Akquisition**
  - Entgegennahme zu archivierender neuer digitaler Inhalte oder Formate
  - Digitalisierung vorhandener physikalischer Archivalien
- **Dokumentation** Katalogisierung der Archivalien. Der Katalog mit den Metadaten bestimmt wesentlich den Nutzwert eines Archivs.
- **Materialsichtung** Kurzfristige, schnelle Bereitstellung von Vorschauen auf die Archivalien. Bündelung von Katalogsuche, Sichtung und Rohschnitt kann den Nutzwert steigern.
- **Wiederverwertung** Auslieferung hochqualitativer Inhalte an Schnittplätze.
- **Asset- und Lebenszyklus-Management** Verwaltung und Steuerung von Prozessen im Lebenszyklus von Medieninhalten, insbesondere Zugriffskontrolle, Versionskontrolle und Rechtemanagement.

Die Umstellung der Fernsehproduktion auf Datei-basierte Plattformen zieht für die Archive neue Anforderungen nach sich [EK05, S. 17]. Sie betreffen insbesondere:

- Hochaufgelöstes Videomaterial (*HighRes*)
- Vorschau des Videomaterials (*LowRes*)
- Audio
- Schlüsselbilder (*Keyframes*)
- Programmbeschreibende Metadaten
- Technische Informationen

Mit dem Wandel stehen die Archive der Rundfunkanstalten vor neuartigen *Herausforderungen* [EK05, S. 17]:

- Strukturierte Verwaltung der Essenz in den neuen Formaten, welche Audio, Video und Daten mitsamt ihrer logischen Struktur verpacken
- Mehrfaches Auftreten der Essenz, z.B. als Original, bitgenaue Kopie, Schlüsselbild (*Keyframe*), Vorschauqualität, usw.
- Verwaltung der technischen Signaleigenschaften der Speicherung
- Verwaltung der strukturellen Eigenschaften der Speicherung
- Verwaltung von programmbeschreibenden Metadaten aus der Abwicklung der Produktion
- Weitere Ergänzungen in den Datenbeständen für einen verlustfreien Metadatenaustausch zwischen Archiv und Produktion, z.B. um Medienbrüche zu vermeiden und Zusammenhänge für Honorar- und Lizenzermittlung aufrecht zu erhalten

Oftmals werden lediglich Ausschnitte der kontinuierlichen Sendeaufzeichnung benötigt. *Partielle Wiederherstellung* ist ein entscheidendes Leistungsmerkmal für das effiziente zur Verfügung stellen von Archivalien.

3.1.6 Digitalisierung der Rundfunkarchivalien

Verbunden mit dem Wandel sind auch bei den öffentlichen Rundfunkanstalten zwei Phänomene zu beobachten:

a) eine ständig anwachsende Menge und Heterogenität von originär in digitaler Form vorliegenden Informationen (*Born Digital*) – In der Akquisition ist die Ablösung des Videobandes durch bandlose Systeme zu erwarten[5].
b) ein zunehmender Umfang von Digitalisierungen von ursprünglich in analoger Form vorliegenden Informationen – In den Archiven ist nach erfolgter Umstellung auf digitale Archive eine Eingliederung früherer Archivalien zu erwarten.

Das *Aussterben von Formaten* ist in physikalischen Medienarchiven ein altbekanntes Problem. Mit der Digitalisierung wird jedoch die Befürchtung verbunden, dass diese Problematik sich verschlimmert [Wri07]. Denn der *Lebenszyklus* eines digitalen Mediums ist im Verhältnis zu herkömmlichen analogen Medien erheblich kürzer.

Digitale Bibliotheken nutzten bereits frühzeitig digitale Technik und haben Verfahren zur Langzeitarchivierung entwickelt [Wri07]. Es werden Hoffnungen darauf gesetzt, dass eine Übertragung dieser Prozesstechnik auf die Medienarchive entscheidende Beiträge zur Bewältigung der neu entstandenen Herausforderungen leisten könnte.

Die ohnehin bereits hohe Quote der Wiederverwendung lässt sich voraussichtlich steigern bei einer konsequenten Weiterentwicklung der Archive zu einem *direkten, öffentlichen Zugriff* auf die Inhalte [Wri07]. Die Digitalisierung bietet hierzu das Potential, denn digitale Archive heben die in physikalischen Archiven gegebene Bindung an den Träger auf und lassen dessen physikalische Handhabung entfallen.

Mit der Digitalisierung ihrer Archivalien verbinden die Rundfunkarchive eine Reihe von *Vorteilen*:

- Effektiver Schutz vor dem Verfall oder gänzlichen Verlust alternder Archivbestände.
- Erheblicher Rückgang der Kosten für die Bestandserhaltung.
- Erhöhte Flexibilität bei der Wiederverwendung der Archivalien.

[5] „Das Vermeiden zeitraubender, nur in Echtzeit ablaufender Handhabungsprozesse […] wird auch im Akquisitionsumfeld die Ablösung des Videobandes durch bandlose Systeme beschleunigen.“ [EK05, S. 15]

Die Umstellung auf digitale Archive ist unvermeidlich [Wri07]. Mit ihnen wird eine Reihe gegenwärtiger Probleme mit physikalischen Trägern, allesamt anzusiedeln im Bereich der Wiederverwendung von Archivalien, gelöst:

- Konkurrenz um die begrenzte Anzahl von Exemplaren einer Medieneinheit
- Umlaufkontrolle
- Einforderung der Rückgabe ausgegebener Medieneinheiten
- Anfertigung zusätzlicher Kopien
- Bereitstellung an weit entfernten Orten

Mit ihnen werden zugleich *neuartige Probleme* auftreten. So erscheint aktuell absehbar das Auftreten andersartiger Fehler. Analoge Medien versagen in der Regel lokal, digitale Medien tendieren bei Versagen zum Totalausfall. Fehlertolerante digitale Medienformate sind bislang ein Wunschtraum [Wri07].

3.1.7 Recherche im Verbund der öffentlich-rechtlichen Rundfunkanstalten

Die in der Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland (ARD) verbundenen Sendeanstalten haben sich vertraglich verpflichtet, einander ihre Medien-Archivbestände zur Verfügung zu stellen.

Mit FESAD betreibt die ARD eine *zentral entwickelte Fernseharchivdatenbank*. FESAD enthält Metadaten von fast allen Produktionen. Das Dokumentationssystem ermöglicht die Recherche nach Bild- und Tonmaterial der Mitgliedsanstalten untereinander. Es befindet sich im Einsatz bei einer ständig wachsenden Zahl von Landesrundfunkanstalten der ARD.

Die Fernseharchivdatenbanken werden dezentral betrieben. Die teilnehmenden Rundfunkanstalten zeichnen selbst dafür verantwortlich, neue Werke und deren Metadaten in den Katalog einzupflegen.

Für einen gemeinsamen Zugriff auf die eigenständigen, verschiedenartigen Archive der angebundenen Rundfunkanstalten wurde in der ARD zudem der Einsatz eines *dezentralen, verteilten Rechechesystems* beschlossen. Mehrere Sendeanstalten nutzen das Vermittlungssystem für die Archivübergreifende Recherche auf den Metadaten der Bestände.

Auch bei dieser Lösung verwenden die Anstalten ihre technischen Systeme weiter. Lokale Erweiterungen übersetzen die Anfragen des Recherchesystems und bedienen auch Anfragen aus anderen Anstalten. Jede nutzende Rundfunkanstalt betreibt eine eigene Instanz des Vermittlungssystems für die Entgegennahme und Verteilung von Rechercheanfragen.

Einige Rundfunkanstalten verwenden des Weiteren ein übergreifendes Recherchesystem, um eine Vielzahl von weiteren Archivkatalogen einzuschließen einschließlich, aber nicht beschränkt auf FESAD, Archimedes, Rechtedaten, Musiksendungen, Wortsendungen, Aussprache, Musik, Geräusche, Noten, Bilddatenbanken, Zeitungen, Agenturmeldungen, Sendeplanung und lieferbare Bücher.

Für einen vereinfachten Informationsaustausch strebt die ARD einen strategischen Rechtedatenverbund an. In dieser Organisationsform werden von den angebundenen Rundfunkanstalten vereinfachte Angaben zu Nutzungsrechten eingeholt und mit den FESAD-Katalogdaten verknüpft. Technisch soll der Verbund alle Rechenanlagen einschließen, auf denen Medienrechte verwaltet werden.

Nicht am Sendezentrum angesiedelte Mitarbeiter recherchieren über Web-Frontends.

3.1.8 Neue Verbreitungswege

Der Rundfunkstaatsvertrag [RStV06, §11(1)] benennt ausdrücklich Hörfunk- und Fernsehprogramme bei der Formulierung des Programmauftrages. Das Bundesverfassungsgericht vertritt darüber die Ansicht, dass dieser Auftrag dynamisch zu interpretieren sei und der öffentlich-rechtliche Rundfunk zu dessen Erfüllung ebenfalls neue Inhalte, Formate und Genres entwickeln können und vor allem neue Verbreitungswege für seine Programme nutzen können muss [BVG07, Absatz-Nr. 123].

Die öffentlich-rechtlichen Rundfunkanstalten unterhalten im Internet bereits eine Reihe von Programmbegleitenden Internetangeboten. Die *Digitalstrategie* des ARD-Verbundes umfasst weit darüber hinaus gehende Pläne für verschiedene digitale Vertriebswege. „Die Palette reicht von HDTV

und Handy-TV über ein Audio- und Videoportal bis hin zu digitalen Zusatzangeboten im Hörfunk“ [ARD07].

Das aktuelle Grundsatzurteil lässt eine Expansion von ARD und ZDF über neue Verbreitungswege erwarten. Öffentlich verwirft die ARD die bisherigen Grenzen für den öffentlichen Rundfunk als „anachronistisch und obsolet“ [HO07]. Mit der kürzlich erfolgten Eröffnung neuer Dienste erweitert der Senderverbund sein Internetangebot. Deutschland folgt damit einer allgemeinen Entwicklung; Im Europäischen Ausland übertragen öffentlich-rechtliche Sender Programminhalte bereits über das Internet, z.B. der ORF.

Als Devise für die deutschen Rundfunkanstalten wird bereits ausgegeben: „[Ihr] Programm muss überall und jederzeit empfangbar sein.“ [Sau07] Eine Vielzahl neuer Möglichkeiten der Verbreitung macht den klassischen Rundfunkwegen Konkurrenz. Gleichzeitig verschwimmen aufgrund der Konvergenz von Fernsehen, Radio und dem Internet die Nutzungsszenarien. Zukünftige Programmformate haben ein nach gleich mehreren Kriterien vielfältigeres *Spektrum* zu bedienen:

- **Verbreitungsweg:** Datennetz (Internet, Mobil)/ Rundfunk
- **Zuschauerverhalten:** konsumierend (lean-back)/ interagierend (lean-forward)
- **Nutzungskontext:** stationär/ portabel-mobil

Bereits heute spielen Rundfunkanstalten ihre Programme auf etwa zehn verschiedene Zielplattformen aus. Neben der terrestrischen, kabelgebundenen und satellitengestützten Ausstrahlung sind dies unter anderem DMB und IPTV.

3.1.9 Menge, Art und Ort anfallender Daten

Daten im Fernsehproduktionsprozess entstammen im Allgemeinen folgender *Herkunft*:

- Sendemitschnitt
- Eigenproduktionen
- Beauftragte Fremdproduktionen
- Eingekaufte oder lizenzierte Produktionen
- Nachrichtenagentur-Meldungen/ -Zusammenschnitte
- Rohmaterial von Eigenproduktionen

Deren Gewichtung im Archivierungsaufkommen ist Rundfunkanstalten-spezifisch.

Die anfallende *Datenmenge* hängt von der Laufzeitlänge der Essenz und dem gewählten Audio-/ Videoformat ab. Eine typische Bitrate für die Archivierung von Audio beträgt 1,5 Mbit/s.

Rundfunkanstalten haben bei der Archivierung von Video ein typisches *Aufkommen* von:

- bis zu 10.000 Stunden im Jahr inklusive Rückwärtsdigitalisierung von Vorbeständen;
- bis zu 8.000 Stunden im Jahr bei Beschränkung auf Vorwärtsdigitalisierung.

Der Aufwand manueller Bearbeitung und Handhabung setzt dem erreichbaren Volumen bei Rückwärtsdigitalisierung eine *Obergrenze*.

Im Mittel umfasst ein Rundfunkarchiv nach zehn Jahren Betrieb 100.000 bis 200.000 Stunden Material.

Die Erfahrungen zeigen, dass im Laufe des Betriebes die bei Neuarchivierung verwendeten *Kompressionsraten* Zug um Zug reduziert werden, so dass bei angenommen gleichem Zeitumfang das *Datenvolumen* steigt. Für die Langzeitarchivierung von Material in Standardauflösung (Standard-Definition, SD) hat sich inzwischen eine Videokompressionsrate von 50 Mbit/s für Programmproduktionen und 25-30 Mbit/s für Nachrichtenproduktionen durchgesetzt.

Hochauflösendes Video (High-Definition, HD) wird sich in Zukunft mittelfristig etablieren. Gegenwärtig wird auch in Deutschland in HD produziert, jedoch nicht in HD gesendet. Breitere Verwendung findet HD bereits jetzt bei Sportproduktionen. Ein einheitliches Format für die Archivierung von HD-Material hat sich zum heutigen Zeitpunkt noch nicht durchgesetzt. Die EBU hat mit 1080i25,

1080p25, 720p50 und 1080p50 vier Systeme vorgeschlagen. Alle diese Formate sind in Europa im Einsatz. Bei der Kodierung ist H.264 (MPEG-4) ein viel versprechender Kandidat. Für die Archivierung stellte HD in erster Linie ein Volumenproblem dar.

3.1.10 Eingesetzte technische Systeme und Art und Umfang der technischen Aufbereitung der Daten

Der Einspielung von Material (*Ingest*) erfolgt in der Regel über die Produktionsumgebung. Erst in einem nachfolgenden Schritt, nicht zwangsläufig unmittelbar anschließend, wird das Material in das Archiv überführt.

Die *Annotation der Essenzen* unterteilt sich in formale und inhaltliche Erschließung.

- **Formale Erschließung** Die formale Erschließung umfasst Mindestangaben wie beispielsweise Titel und Sendezeit und ist strengen Richtlinien unterworfen.
- **Inhaltliche Erschließung** Die inhaltliche Erschließung ist dem gegenüber subjektiv geprägt in ihrer Form, ihrem Umfang und ihrer Wortwahl. Die Ausrichtung und Zielsetzung der Rundfunkanstalt beeinflusst maßgeblich, inwieweit Sachinhalte, Bildinhalte und Rechte bei der inhaltlichen Erschließung eingearbeitet werden.

Die Erschließung wird maßgeblich manuell durch Dokumentare geleistet. Sie werden unterstützt durch Texterkennungssysteme (OCR). Sprach- und Sprechererkennung sind allenfalls experimentell im Einsatz und mit einer hohen Fehlerquote behaftet. Da die Dokumentation das Ziel hat, einen hohen *Wiederverwertungsgrad* zu erzielen, steht die Qualität der Suchergebnisse (*Precision*[6]) im Vordergrund. Daher ist bei automatischen Indizierungswerkzeugen an die Fehlerfreiheit der Ergebnisse ein besonders hoher Anspruch zu stellen.

3.1.11 Technische Infrastruktur

Die vernetzte Produktionsumgebung (Abbildung 12) lässt sich gliedern anhand der organisatorischen Bereiche der Rundfunkanstalt: Redaktion, Produktion-&-Postproduktion, Ingest, Portal, Sendeabwicklung und Archiv. Innerhalb der Bereiche kommen auf die jeweiligen Aufgaben spezialisierte Geräte und Anwendungen zum Einsatz. Geräte als auch Bereiche sind über Informationstechnik-Netzwerke miteinander verbunden.

[6] Im Information-Retrieval (IR) bezeichnet Precision den Anteil tatsächlich relevanter Dokumente in der Gesamtmenge bei einer Suchanfrage ermittelter Dokumente.

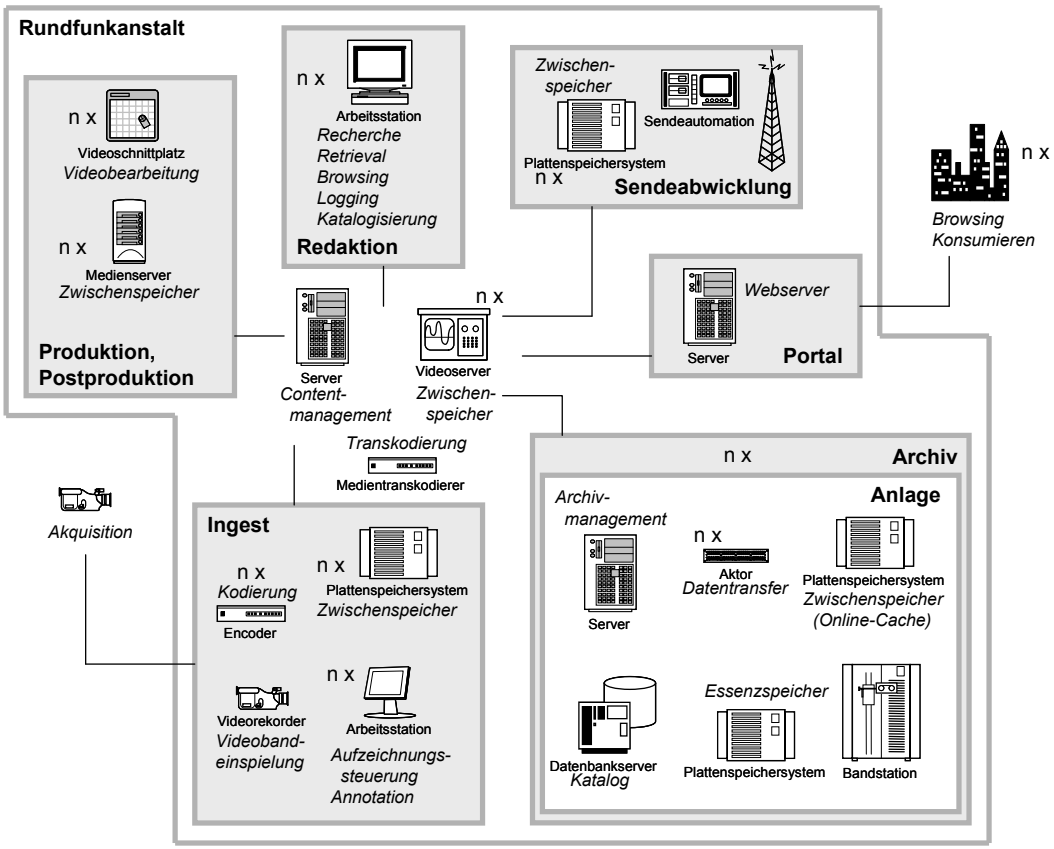

Abbildung 12: Grundsätzliche Komponenten sowie deren Funktion der allgemeinen technischen Infrastruktur der Rundfunkanstalten und deren organisatorische Gliederung bei Datei-basierter Fernsehproduktion, vgl. [EK05, S. 16].

Typischerweise kommen folgende *technische Systeme* in der Produktions- und Archivumgebung von Rundfunkanstalten zum Einsatz.

Hardware:

- **Bandbibliotheken (*Tape Libraries*), Bandlaufwerke und Bandroboter** Bandbibliotheken finden in den Rundfunkanstalten hauptsächlich für die *dauerhafte Speicherung und Archivierung der Essenzen* Verwendung. Sie verwahren zudem die *Datensicherung* (*Backup*) von Katalog- und Metadaten.
- **Plattenspeichersysteme** Das klassische Einsatzfeld von Festplattensystemen ist die *Zwischenspeicherung* (*Nearline/ Online-Cache*) von Essenz für eine performante Bereitstellung. Zunehmend etablieren sich Festplattensysteme als Alternative zu Bandbibliotheken als *dauerhafter Archivspeicher*. In dieser Funktion lagern sie bereits Metadaten sowie Essenzdaten in Vorschauqualität und Audio in Produktionsqualität.
- **Aktoren/ Datenpumpen (*Actors/ Data Movers*)** Datenpumpen sind für den *Transport* in der Regel umfangreicher Datenbestände zwischen Speichersystemen zuständig. Sie vermitteln bei dem Datendurchsatz zwischen Bandbreiten und Protokollen verschiedener Netztechniken als auch Speichersystemen – insbesondere Band und Festplatte. Indem sie Transaktionen eigenständig abwickeln entlasten sie den Anwendungsserver des Datenmanagements.
- **Medienserver** Medienserver vereinen mit Essenzspeicherung, En- und Dekodierung, Videobearbeitung und Contentverwaltung grundlegende Funktionalitäten der digitalen, Datei-basierten Produktion.
- **Verteilte Medienspeicher** Verteilte Medienspeicher sind Plattenspeichersysteme mit einem für den Medienbereich erweiterten Leistungsumfang. Sie zeichnen sich durch Skalierbarkeit, Bandbreite, Verfügbarkeit, Echtzeitbereitstellung und Parallelzugriff aus.

- **Anwendungsserver (*Application Server*)** Diese regulären Computersysteme beheimaten die Vielzahl von Softwaresystemen zur Verwaltung und Steuerung des Archivs als auch des Produktionsbetriebs.
- **Netzwerk (*Network*)** Informationstechnik-Netzwerke erschließen durchgängig die Rundfunkanstalten (Intranet/ Local Area Network, LAN). Hierbei überwiegt 1 Gbit/s-Ethernet in Abteilungen und Etagen sowie 100 Mbit/s-Ethernet zum einzelnen Arbeitsplatz mit Protokoll TCP/IP. Absehbar ist eine Steigerung auf 10 Gbit/s. Innerhalb der Archivabteilungen werden zudem Speichernetzwerke (Storage Area Network, SAN) mit Fibre Channel (FC)-Protokoll eingesetzt.
- **Hardware-Kodierer (*Encoder*), -Dekodierer (*Decoder*)** Diese spezialisierten Komponenten *wandeln Audio- und Videosignale in digitale Formate* respektive umgekehrt. Sie vermitteln zwischen klassischer Audio- und Videotechnik einerseits und Datei-basierter Produktion und Archivierung andererseits.
- **Arbeitsplatzrechner (*Personal Computers*)** Reguläre Computersysteme dienen in den Rundfunkanstalten als *universelle Endgeräte* für nahezu sämtliche Aufgaben. Für anspruchsvolle Arbeiten wie Videobearbeitung kommen leistungsfähige Arbeitsplatzrechner zum Einsatz, dann auch ergänzt um zusätzliche Ein- und Ausgabekomponenten.
- **Sendeautomation** Eine Sendeautomation erlaubt anhand zeitlicher Transmissionslisten die Planung und automatische, unterbrechungsfreie Ausspielung von Audio- und Videomaterial.
- **Videoserver** Ein Videoserver ist in der Lage, mehrere Videoströme gleichzeitig und mit fremder Synchronisation innerhalb der Spezifikation für Videosignale auszuspielen.

Mit steigender Leistungsfähigkeit wird zunehmend angepasste *Standard-Informationstechnik* anstatt spezieller teuerer Gerätetechnik eingesetzt [Röd07].

Software:

- **Archivmanagement (*Archive Management*)** Das Archiv-Management organisiert die Speicherung von Inhalten und deren Bewahrung auf den verschiedenartigen Speichersystemen des Archivs – in der Regel Band und Festplatte. Weiterhin steuert es die *Archivierungs- und Wiederherstellungsprozesse* zur Einlagerung von Produktionsergebnissen respektive der Bereitstellung von Archivalien für Produktion und Sendeabwicklung. Dies schließt die Vervielfältigung und Spiegelung von Archivinhalten mit ein.
- **Dateitransfer** Werkzeuge zum Dateitransfer wickeln die *Verlagerung oder das Erstellen einer Kopie von Essenz zwischen Serversystemen* ab. Zu ihren Leistungsmerkmalen gehören Bandbreitenbündelung, Übertragungswegredundanz und Lastverteilung.
- **Media-Asset-Managementsysteme (*MAM*)** Media-Asset-Managementsysteme verwalten die Essenzen einer Rundfunkanstalt anhand von Metadaten, und bieten optional die Verwaltung der formalen und inhaltsbeschreibenden Metadaten. Sie erlauben die *Recherche* im Essenzbestand und vermitteln anhand von technischen, formalen und inhaltsbeschreibenden Metadaten, Schlüsselbildern (*Keyframes*), Vorschaufassungen (*LowRes*) sowie Begleitdokumenten (*Collaterals*).
- **Datenbanken (*Databases*)** Relationale Datenbanken führen die *Essenz-Metadaten* sowie *Verwaltungsdaten* von Media-Asset-Management (*MAM*), Datenmanagement und Archivmanagement.
- **Transkodierer (Transcoder)** Transkodierer *wandeln die Kodierungs- und Dateiformate digitaler Essenz*.
- **Qualitätswerkzeuge** Diese automatisierten oder halb-automatisierten Verfahren prüfen die Qualität von Videoinformation, beispielsweise durch Schwarzbilderkennung u ä.

Hohe Verfügbarkeit und Realzeitfähigkeit sind die wesentlichen Merkmale der IT-Lösungen für den Rundfunk. Aufrüstungen von Hard- und Software müssen im laufenden Betrieb erfolgen. Die ausgewählten Lösungen müssen Anpassungen erlauben. Um eine hohe Ausfallsicherheit zu garantieren, sind nahezu alle Systemkomponenten der Archivlösungen *redundant* ausgelegt.

Aus der Erfahrung steht die Anzahl der registrierten Benutzer bzw. der Arbeitsplatzrechner, von denen aus auf das System zugegriffen werden kann, in einem Verhältnis von 3:1 zu der Anzahl der tatsächlichen gleichzeitigen Benutzer.

In der Regel bieten die öffentlich-rechtlichen Rundfunkanstalten in Deutschland allerdings *keine* öffentliche Schnittstelle für einen Archivzugang über das Internet an.

3.1.12 Technischen Plattformen, deren Eigenschaften und Mechanismen zur Speicherung von hochkomplexen und verteilt liegenden Inhalten

Verteilte Speicherung in der vernetzten Produktionsumgebung einer Rundfunkanstalt

Gemeinsam genutzter Speicher ist ein Schlüsselmerkmal der Architektur vernetzter Produktionsumgebungen in Rundfunkanstalten [WG06]. Dessen grundsätzlicher Aufbau folgt in vielerlei Hinsicht der hierarchischen Organisation zentral organisierter Speichersysteme (s. Kapitel 2). In der technischen Infrastruktur der Rundfunkanstalten sind typischerweise eine Reihe von Speichersystemen anzufinden (Abbildung 12), welche sich in Online-, Nearline-, Archiv- und Offlinespeicher einteilen lassen:

- In Produktion und Postproduktion stellen Medienserver den Videoschnittplätzen das Material bereit (Online).
- In der Sendeabwicklung nehmen Sendespeicher frühzeitig von der Sendeautomation für die Ausstrahlung vorgesehene Beiträge auf (Online).
- Im Ingest nehmen Medienserver neu digitalisiertes Material auf und halten es für die formale und inhaltliche Erschließung bereit (Online).
- Videoserver oder Verteilte-Medienspeicher speichern als Zentralspeicher in der Produktion kürzlich genutztes Material (Nearline).
- Im Archiv verwahren Bandbibliotheken und Plattenspeicher Material langfristig (Archiv). Plattenspeicher speichern bei Archivzugriffen Material zwischen (Nearline).

Innerhalb eines Lebenszyklus durchlaufen Inhalte die verschiedenen Speichersysteme. Essenzen können in mehreren Kopien oder auch als Kopien an mehreren Orten zugleich angelegt werden. Die Regeln, nach denen Material zwischen den Speichersystemen bewegt wird, sind abhängig von den Inhalten und den Geschäftsprozessen und damit hochgradig komplex. Das Content-Management verwaltet, im Allgemeinen in einer Datenbank, für den Geltungsrahmen Rundfunkanstalt den Speicherort von Material und dessen Kopien. Anhand von Verwaltungs- und Indexinformation findet es Material, fordert das speichernde System auf es wiederherzustellen und löst ein Kopieren oder Verschieben an die benötigte Stelle aus. Einlieferung und Auslieferung von Essenz und Metadaten in das Archiv respektive aus dem Archiv sind Bestandteil dieser Architektur.

Der Ansatz, für die Bereiche der Rundfunkanstalt das jeweils beste System seiner Art einzusetzen führt zu einer heterogenen Infrastruktur. Die Teilsysteme stammen von verschiedenen Herstellern und, zu nennen sind hier vor allem Verteilte-Medienspeicher, Medienserver und Videoserver, führen ein eigenes Speichermanagement. Ihre Integration zu einem gemeinsamen Speicher macht deren Synchronisation erforderlich.

Die Auslegung dieser technischen Infrastruktur hat sich an den individuellen Zugriffsmustern der jeweiligen Installation zu orientieren. Diese bestimmen sich anhand der Arbeitsprozesse in der jeweiligen Rundfunkanstalt.

Verteilte Speicherung in einzelnen Speichersystemen

Einzelne Speichersysteme weisen in sich eine *Clusterarchitektur* auf [WG06]. Sie bestehen aus einer Anzahl von Knoten mit Speicher-, Rechen- und Kommunikationskapazität. In einer typischen Konfiguration stellt ein redundant ausgelegter Datenbank-Knoten die Schnittstelle für die Anwendungen bereit und weist den Speicherknoten Aufgaben zu. Wesentliche Leistungsmerkmale der

Clusterverwaltung sind Lastverteilung, Auftragspriorisierung und Ressourcenplanung. Den Datentransfer innerhalb des Clusters und zu den Anwendungssystemen wickeln die Speicherknoten eigenständig ab. Sie kommunizieren über Fiberchannel oder Ethernet als *Network Attached Storage (NAS)* respektive Storage *Area Network (SAN)*. Regelmäßige Kontrollsignale (*Heartbeat*) lassen den Verwaltungsknoten Knotenausfälle erkennen. In einer solchen Situation kann ein anderer Knoten die Aufgaben des ausgefallenen Knoten übernehmen. Im Falle eines wesentlichen Dienstes erfolgt diese Umorganisation automatisch.

Verteilte Speicherung über Produktionsstandorte hinaus

Rundfunkanstalten betreiben gelegentlich mehrere Hauptstandorte sowie in der Regel eine Reihe von Regionalstudios und Auslandsbüros. Innerhalb Deutschlands sind die Standorte von Rundfunkanstalten im Regelfall mit Standleitungen untereinander verbunden. Im Ausland werden derartige Verbindungen durchaus auch dynamisch aufgebaut.

Die *Replikation* von verwalteten Assets ist kein grundsätzliches Problem. Die in einem Verbund verwalteten Inhalte können miteinander synchronisiert werden. Die Eigentümerschaft an replizierten Inhalten kann von einem auf ein anderes System übergehen. In der Regel wird dies zentralisiert verwaltet. Stand der Technik ist eine *Tokenregelung* zur Kennzeichnung, welcher Standort die Hoheit über die Referenzkopie innehat. Kritisch hingegen sind massive, verteilte, synchrone Transaktionen über viele Standorte.

Verteilte Speicherung auf Datenbändern

Bei der Speicherung in Bandbibliotheken werden Dateien ohne Unterbrechung fortlaufend auf *Datenband* abgelegt. Eine Verteilung über mehrere Bänder ist unerwünscht, aber möglich und aus Kapazitätsgründen teils unvermeidlich. Insbesondere erfolgt keine absichtliche Streuung von Essenz über Datenbänder hinweg mit dem Ziel der Schaffung von Sicherheit durch Redundanz oder Lastverteilung. Der Datenträger Datenband erzielt den höchsten Durchsatz bei kontinuierlichen Lese-/Schreibvorgängen. Bandbereitstellung und Bandpositionierung haben wesentlichen Anteil an der Gesamtzugriffszeit. Daher würde eine Verteilung über mehrere Bänder im Regelfall keinen positiven Beitrag zu Verbesserung der Performance leisten und wäre, mit Blick auf mögliche maximale Anzahl gleichzeitiger Dateiübertragungsvorgänge, durch die pro Transfervorgang höhere Anzahl von Bandlaufwerken sogar schädlich

Technisch bedingt erfolgt bei der *Löschung* einer auf Datenband gespeicherten Essenz zunächst lediglich ein Löschungsvermerk im Bandinhaltsverzeichnis. Erst später, entweder beim Migrieren auf ein neues Datenband oder beim bewusst angestoßenen Defragmentieren, werden die Essenzdaten tatsächlich gelöscht.

3.1.13 Inhaltsbeschreibungen und Rechte-Daten sowie Metadaten

Daten in der Fernsehproduktion

In der Fernsehproduktion werden jene Daten (Abbildung 13) als *Essenz* bezeichnet, welche den primären Inhalt des Medienproduktes darstellen [Röd07]. *Metadaten* enthalten Information über die Essenz. Sie sind „für die effiziente Planung zur Erstellung und weitgehend automatisierten Bearbeitung der Nutzdaten (Video, Audio und Zusatzdaten), deren Verwaltung und Speicherung sowie deren Austausch erforderlich“ [Ebn05]. Im Gegensatz zu *Zusatzdaten* sind Metadaten ohne die durch sie beschriebene Essenz ohne Wert [Röd07]. *Content* bezeichnet die Verbindung aus Essenz und Metadaten. Um Rechtedaten ergänzt wird Content zum *Asset*.

Content-Management bezeichnet die Verwaltung von Essenz mittels (formal) beschreibender Metadaten. Das *Asset-Management* berücksichtigt darüber hinaus Rechte [Sau02].

Als *Datenarten* werden in den Rundfunkarchiven unterschieden:

- Essenz in **Produktionsqualität** (*HighRes*)
- Essenz in **Vorschauqualität** (*LowRes*)
- **Metadaten**

Diese Unterscheidung betont das Volumen und die Zugriffshäufigkeit der Daten. Diese Kriterien beeinflussen entscheidend die Gestaltung der technischen Infrastruktur von Archiv und Produktionsumgebung.

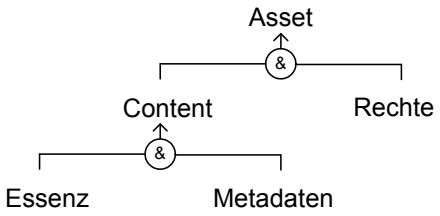

Abbildung 13: Daten in der Fernsehproduktion: Essenz, Metadaten, Content, Rechte, Asset.

Metadaten

Metadaten können unter anderem nach ihrer *Notwendigkeit* kategorisiert werden.

- **Technische Metadaten** sind eine Voraussetzung für die korrekte Dekodierung der Essenz. Sie sind relevant für die *Anwendung* der Essenz und deren Darstellung.
- **Beschreibende Metadaten** sind nicht grundsätzlich erforderlich, aber für übergeordnete Belange wie die Organisation des Produktionsprozesses und die Distribution hilfreich. Sie sind relevant für das *Auffinden* der Essenz.

Die Erstellung, Änderung und Verwaltung dieser Metadaten wird Bestandteil des Produktionsprozesses und elektronisch erfolgen [Sau07]. Während der Produktion werden umfassend Metadaten mit bis zu 500 Attributen über die Essenz geführt. Das Langzeitarchiv enthält dem gegenüber deutlich weniger Metadaten in der Größenordnung von 30 Attributen. Über den erforderlichen Satz von Metadaten zur Produktion hinaus entscheidet die Qualität der Metadaten über die Präsenz des Materials beim Konsumenten [Sau07].

„Es gibt keinen abgestimmten und weitgehend eingeführten Standard für die Speicherkomponenten in Assetmanagement-, Schnitt- und Playoutsystemen. Es ist das Rundfunkarchiv, das ein allgemeines Format für Metadaten und Media vorgibt.“ [Wri07]

In der Regel verwenden keine zwei Rundfunkanstalten das gleiche Metadatenmodell. Eine Ausnahme sind die in der ARD organisierten Anstalten, von denen die überwiegende Zahl im FESAD-Konsortium organisiert ist und das von FESADneu angebotene Datenmodell verwenden.

Bei der Einführung eines Asset-Managementsystems bestehen drei alternative Ansätze für den *Aufbau des Metadatenschemas*:

- Weiterverwendung der in der Regel vorbestehenden Datenbank mitsamt ihres Datenschemas
- Erarbeitung eines Datenschemavorschlags durch den Anbieter/ Entwickler des Asset-Managementsystems
- Verwendung eines Standarddatenschemas

Anbieter von Asset-Managementsystemen setzen derzeit *BMF* erstmalig als Metadatenschema um.

In einer geschlossenen Umgebung kann die Aufbewahrung und Bereitstellung von Metadaten zweierlei organisiert werden [Röd07]:

- **Zentrale Speicherung** Die Metadaten werden, getrennt von der Essenz, in einem zentralen Katalog geführt. Für diesen Ansatz spricht eine hohe Sucheffizienz, weil Metadaten nicht erst aus der Essenz geborgen werden muss. Voraussetzung sind zuverlässige Verknüpfungsmechanismen.
- **Dezentrale Speicherung** Die Metadaten werden in ihrer Essenz eingebettet. Mit diesem Ansatz stehen die Metadaten unabhängig von einer zentralen Instanz wie dem Katalog bereit.

Eine Kombination beider Organisationsformen resultiert potentiell in *Inkonsistenzen* und Abgleichungsproblemen zwischen abweichenden Metadatenangaben.

Rechtedaten in der Fernsehproduktion

Die Handhabung und Berücksichtigung von Rechten an Ton- und Bildmaterial ist Bestandteil des Fernsehproduktionsprozesses. In der ARD verwalten die Abteilungen für Honorare und Lizenzen (HoLi) diese Verträge. Historisch haben sich Rechte- und Medienarchive zum großen Teil unabhängig voneinander entwickelt [Sau02]. Die Verwaltung von Rechtedaten zu den Beständen in einer Sendeanstalt ist derzeit Aufgabe eigenständiger, regelmäßig Datenbank-basierter Systeme, getrennt vom eigentlichen Bild- und Tonmaterial [NN01]. Die Archive der Sendeanstalten sind an diese Rechteinformationssysteme angeschlossen, in der Regel zu Zwecken der Archiverfassung. Nicht in jedem Fall setzen die Archive die sich aus den Rechtedaten ergebenden Einschränkungen und Vorschriften automatisiert um. Eine Recherche über Archivkatalog und Rechteinformationssystem zugleich war nur bedingt – wenn überhaupt – möglich, weshalb Rechteinformation oftmals erst zu spät Berücksichtigung fand [Sau02].

Es wird erwartet, dass in digitalisierten Produktionsumgebungen die Rechteverwaltung durch Verwendung von Archiven und Content-Managementsystemen samt zugehöriger Metadaten problemlos umsetzbar sein wird [Sau07].

Die Verwaltung von Rechten an Essenz ist hochgradig komplex. Nutzungsverträge werden in der Regel individuell ausgehandelt und enthalten Sonderregelungen. Ebenfalls schwierig zu verwalten sind dynamische Rechte, beispielsweise Senderecht ab drei Stunden nach erstmaliger Ausstrahlung durch den Inhaber der primären Senderechte. In Summe ist die Rechteverwaltung *nicht vollständig schematisierbar*. Zu Teilen lassen sich Nutzungsvereinbarungen lediglich prozedural formulieren. Für Asset-Managementsysteme ist dieser Grad der Rechteverwaltung oftmals zu komplex.

Es haben sich daher spezialisierte, kommerzielle Lösungen für die Rechteverwaltung herausgebildet. Auch setzen die öffentlich-rechtlichen Rundfunkanstalten für diese Aufgabe Eigenentwicklungen ein. Die installierten technischen Lösungen im Bereich Honorare-&-Lizenzen in den öffentlich-rechtlichen Rundfunkanstalten sind verschiedenartig. [NN01]

Für die Verwaltung und Abgeltung von Urheber- und Nutzungsrechten an Essenz führen die Abteilungen Honorare-&-Lizenzen nachfolgende Daten: [NN01]

- Angaben zum Rechteinhaber
- Angaben zu den Inhalten
- Angaben zum Nutzer
- Angaben zu den zulässigen Nutzungsformen
- Angaben zur vereinbarten Vergütung
- Angaben zu erfolgter Nutzung (Ausstrahlung)
- Angaben zu erfolgter Vergütung

Wie bei Metadaten allgemein ist auch bei den Rechtedaten ein Trend hin zu stark feinerer Untergliederung zu beobachten.

Eigenproduktionen machen bei den öffentlich-rechtlichen Rundfunkanstalten einen Anteil im Bereich von 2/3 bis 3/4 aller Beiträge aus. Die zugekauften Inhalte sind also in der Minderzahl.

Die Verwaltung von Rechtedaten innerhalb der Rundfunkanstalt, insbesondere in Produktion und Archiv, ist abzugrenzen gegenüber dem *Digitalen Rechtemanagement (DRM)*. Letzteres bezeichnet Verfahren zur Kontrolle der Nutzung und Verbreitung digitaler Medien. Rechteinhaber vergeben zu von ihnen festgelegten Konditionen Nutzungsrechte an Inhalten. Formale Beschreibungen dieser Nutzungsrechte werden an die Essenz gebunden. Digital-Rights-Management-Systeme (DRMS) erzwingen technisch die Einhaltung dieser vereinbarten Nutzungsbedingungen. Zum Schutz vor unbefugtem Zugriff basiert DRM auf Verschlüsselungsverfahren.

Am Digitalen Rechtemanagement interessiert sind in Zukunft voraussichtlich die Hersteller von Inhalten, insbesondere Filmindustrie, Bezahlfernsehen-Anbieter und Sendeanstalten. Die deutschen öffentlich-rechtlichen Rundfunkanstalten verfolgen für ihre Sendungen das *Modell des freien Empfangs* und haben sich stets gegen eine Verschlüsselung ausgesprochen.

Innerhalb der öffentlich-rechtlichen Rundfunkanstalten regeln *organisatorische Maßnahmen* die Einhaltung von Nutzungsrechten an Essenzen. Essenz in Produktion und Archiv ist frei von einem technischen Schutz durch Digitales Rechtemanagement. Die Produktionsgeräte sind grundsätzlich nicht für einen Umgang mit DRM ausgelegt. Allenfalls, dies betrifft Vorschaumaterial, wird Essenz mit Digitalen Wasserzeichen (*Watermarking*) gekennzeichnet.

Essenzdatenformate

Für den Datei-basierten *Austausch von Essenz und Metadaten im Produktionsprozess* stehen diverse Formate bereit [Röd07]. Als standardisierte Formate zu nennen sind hier **DPX** (Digital Moving Picture Exchange, SMPTE268M), **GXF** (General Exchange Format, SMPTE360M) und **MXF** (Material Exchange Format, SMPTE377M ff.). **AAF** (Advanced Authoring Format) ist nicht standardisiert, genießt aber eine hohe Akzeptanz in der Industrie. In der Fernsehproduktion bedienen die Formate die unterschiedlichen Anwendungsbereiche Filmabtastung und Rendering, Übertragung kompletter Programmteile sowie Postproduktion. In der Behandlung des Materials und den Metadaten gehen die Formate verschiedene Wege. Während GXF das Bildmaterial komprimiert, kodiert DPX unkomprimiert; MXF als auch AAF können mit beiderlei Verfahren umgehen. Der Umfang der übertragbaren Metadaten reicht von Strukturangaben und beschreibenden Metadaten bis hin zu einer Beschreibung kompletter Bearbeitungsschritte oder überlässt es dem Anwender, eigene Datenbereiche zu definieren.

Das Containerformat **MXF** wird von Herstellern sowie Anwendern zunehmend anerkannt.

Metadatenformate

Gegenwärtig ist eine Vielzahl von proprietären Datenschnittstellen im Einsatz und auch zukünftig wird eine Vielzahl von Formaten zu berücksichtigen sein. „Erschwerend kommt hinzu, dass es bis heute kein gemeinsames Verständnis für die gleiche Information [sic] [in den Metadaten] […] gibt." [EK05]

Über zahlreiche *Eigenentwicklungen* der Rundfunkanstalten und proprietäre *Herstellerformate* hinaus lassen sich einige wenige Metadatenschemata ausmachen.

BMF[7] Das Broadcast Metadata Exchange Format (BMF) wurde vom Institut für Rundfunktechnik (IRT) entwickelt mit dem Ziel, ein einheitliches, generisches, herstellerunabhängiges Datenmodell für sämtliche Metadaten in der Fernsehproduktion bereitzustellen, welches maximale Interoperabilität zwischen allen Produktionsbereichen und Rundfunkanstalten ermöglicht. Das Format erhebt den Anspruch, nahezu alle Bereiche im Lebenszyklus eines Fernseh-Programmbeitrags abzudecken. Das Datenmodell berücksichtigt sowohl redaktionelle Konzepte als auch eine Beschreibung der Signale und deren Speicherung.

FESAD FESADneu ist das neue Datenbankmodell für die Datenbank der ARD-Fernseharchive [Ebn05]. Zusätzlich zu den strukturierten Attributen erlauben Freitextfelder beliebige Angaben, welche im praktischen Einsatz durch Volltextsuchen ebenfalls - auch automatisiert - ausgewertet werden. Rechtedaten sind nicht Bestandteil.

PBCore Das Public Broadcasting Metadata Dictionary (PBCore) [CPB07] ist ein Metadatenschema zur Beschreibung und Kategorisierung von Medieneinheiten, ausgerichtet auf den Amerikanischen Rundfunk und verwandte Branchen. Seine Entwicklung wird gefördert von der Corporation for Public Broadcasting (CPB), einer öffentlich finanzierten Organisation in den Vereinigten Staaten. Seine 53 Attribute behandeln die Inhalt, Urheber- und Nutzungsrechte als auch Format und lassen Raum für Erweiterungen einzelner Nutzergruppen. Das Modell basiert auf *Dublin Core*, einem etablierten internationalen Standard (ISO 15836) zur Beschreibung von Dokumenten und anderen Objekten im Internet.

DMS1 Das Descriptive Metadata Schema (DMS1) ist ein standardisiertes Format der SMPTE zum Austausch von Metadaten in Verbindung mit MXF. Es ist umfassend in der Beschreibung auszutauschenden Materials, unterstützt jedoch nicht ganze Produktionsprozesse.

[7] http://www.irt.de/de/produkte/produktion/bmf.html

3.2 Hochschul-Medienzentren

Digitale Langzeitarchivierung wird innerhalb von Hochschul-Medienzentren zur Bereitstellung von Information von öffentlichem Interesse eingesetzt. Zum einen wird ursprünglich analog vorliegende Information digitalisiert und so vor zerstörenden Alterungsprozessen geschützt. Zum anderen wird originär in digitaler Form entstandene (*Born Digital*) Information in ein Archiv aufgenommen und erhalten. Dazu zählen elektronische Publikationen ebenso wie Multimediainhalte aus dem Bereich e-Learning.

3.2.1 Digitalisierung der Information für Hochschul-Medienzentren

Verbunden mit dem Wandel sind bei den Hochschul-Medienzentren zwei Phänomene zu beobachten:

a) ein zunehmender Umfang von Digitalisierungen von ursprünglich in analoger Form vorliegender Information
   - In den Archiven ist nach erfolgter Umstellung auf digitale Systeme eine Eingliederung früherer Archivalien zu erwarten.
   - In der Akquisition/ Sammlung ist die Ablösung alter Papierbestände zu erwarten.

b) eine ständig anwachsende Menge und Heterogenität von originär in digitaler Form vorliegender Information (*Born Digital*).

Bisherige *analoge Formate* erreichen zunehmend ihre Altersgrenze und sind von dem physikalischen Zerfall bedroht. Mit der Digitalisierung ist hier zunächst eine kurzfristige Lösung gefunden. Langfristig ist die Problematik der Alterung dadurch verschlimmert, denn der *Lebenszyklus* eines digitalen Mediums ist im Verhältnis zu herkömmlichen analogen Medien, wie beispielsweise Papier, erheblich kürzer. Demnach müssen Verfahren zur digitalen Langzeitarchivierung entwickelt werden, die Archivierung nicht als statischen Zustand behandeln, sondern als dynamischen Prozess ansehen. Darüber hinaus muss ein digitales Langzeitarchiv vertrauenswürdig und abgesichert sein, Anforderungen, wie sie eingangs aufgezeigt wurden, die mit der Einführung der digitalen Langzeitarchivierung eine neue Bedeutung bekommen haben. Denn sie stellen gleichzeitig die neuen Herausforderungen dar.

Mit der Digitalisierung von Information verbinden die Hochschul-Medienzentren eine Reihe von *Vorteilen*:

- Effektiver Schutz vor dem Verfall oder gänzlichen Verlust alternder analoger Archivbestände
- Verringerung der Kosten und Erhöhung der Wirtschaftlichkeit
- Erhöhte Flexibilität (orts- und zeitunabhängig) bei der Wiederverwendung der Archivalien
- Effizienter Umgang mit digitalen Archivalien
  - Keine begrenzte Anzahl der Exemplare
  - Umlaufkontrolle
  - Einfache Kopiengenerierung und Auslieferung/ Zugriff

Die Umstellung auf digitale Langzeitarchive ist gleichzeitig aber auch mit *Nachteilen* und neuen Gefahren verbunden. Zu den aktuell absehbaren Problemen zählen:

- Trennung von Speicherung und Darstellung
- Bereitstellung geeigneter Abspielsysteme
- Technikausfall
- Aktualität
- Verlust der Information („Schwarzes Loch“)
- Manipulationen

3.2.2 Gefährdungen gemäß BSI

An dieser Stelle sei noch einmal die *Gefährdungskategorisierung* des Bundesamts für Sicherheit in der Informationstechnik (BSI)[8] in Bezug auf digitale Langzeitarchivierung aufgeführt, da diese die absehbaren Probleme beinhaltet:

G 1: Höhere Gewalt

G 2: Organisatorische Mängel

G 3: Menschliche Fehlhandlungen

G 4: Technisches Versagen

G 5: Vorsätzliche Handlungen

Die Organisation und Umsetzung digitaler Langzeitarchivierungssysteme ist darauf bedacht, diese Gefährdungen zu berücksichtigen, um vertrauenswürdig und abgesichert zu sein.

3.2.3 Digitale Langzeitarchivierung in Hochschul-Medienzentren – Auftrag und generelle Aufgaben

Die digitale Langzeitarchivierung dient der Erhaltung und Bereitstellung unserer kulturellen Überlieferungen sowie ihrer dauerhaften Verfügbarkeit für Wissenschaft und Forschung [ScLi04]. Digitale Ressourcen sollen langfristig verfügbar gehalten werden. Dies ist unter anderem dadurch zu begründen, dass im *Wissenschaftsprozess* langfristig archivierte Fakten immer wieder neu bewertet werden und daher zugänglich sein müssen. Die Langzeitarchivierung digitaler Ressourcen ist daher eine wesentliche Bedingung für die Konkurrenzfähigkeit des Bildungs- und Wissenschaftssystems und der Wirtschaft [ScLi04]. In diesem Auftrag sind digitale Langzeitarchive in Hochschul-Medienzentren eingesetzt. Publikationen sollen so veröffentlicht, verbreitet und erhalten werden. Schwerpunkt dabei ist zum einen die Substanzerhaltung der Information und die Erhaltung der Benutzbarkeit, wobei die Erhaltung der Benutzbarkeit nicht unbedingt vereinbar ist mit der Erhaltung der ursprünglichen Ausprägung des originalen Informationsobjekts. So werden Transformationen notwendig sein, welche die ursprüngliche Ausprägung eines Informationsobjekts verändern, um seine Benutzbarkeit zu garantieren.

3.2.4 Akteure

Für Langzeitarchivierungssysteme in Hochschul-Medienzentren gibt es unterschiedliche *Akteure*:

- Hochschule
- Zentrale Einrichtungen wie
  - Bibliotheken
  - Rechenzentren
- Archivdienstleister
- Institut
- Mitarbeiter
- Studenten
- Externe Personen

Diese Akteure haben verschiedene und dabei gleichzeitige Aufgaben- und Wirkungsbereiche, die sich unterteilen in archivintern und archivextern. Der *archivinterne Bereich* umfasst Aufgaben der Verwaltung und Bereitstellung des Archivs, der Archivumgebung, der technischen Infrastruktur und der Archivobjekte. Hierzu zählen die bereits im nestor-Kriterienkatalog aufgeführten organisatorischen, technischen und rechtlichen Verantwortungen. Solche Verantwortungen liegen grundsätzlich bei der das Archiv bereitstellenden Instanz. In diesem Fall ist es die Hochschule, die das Langzeitarchiv anbietet. Die Hochschule gibt die Verantwortlichkeiten weiter und verteilt und Aufgaben an die durch-

[8] Weiterführende Ausführungen siehe http://www.bsi.de.

führenden Institutionen, die der Hochschule angehören. Dies sind in der Regel zentrale Einrichtungen wie Rechenzentren und Bibliotheken. Die Hochschule kann Verantwortungen und Aufgabenbereiche aber auch auslagern an Dritte, externe Archivdienstleister, die der Hochschule nicht angehören. Die archivinternen Aufgabenbereiche beinhalten ebenso die Beachtung von Richtlinien und Rechten.

Der *archivexterne Bereich* umfasst Aufgaben- und Wirkungsbereiche außenstehender Parteien und externer Personen, die zu archivierende Objekte erstellen und einliefern (Produzent) oder auf archivierte Objekte zugreifen (Konsument). Hier ist festzuhalten, dass Produzenten gleichzeitig auch Konsumenten sein können. Produzenten sind beispielsweise Mitarbeiter der Hochschule, die auch gleichzeitig Konsumenten sein können. Konsumenten sind aber auch Studenten oder externe Personen, die auf das Archiv zugreifen, um an archivierte Information zu gelangen. Im Falle von e-Learning ist ein Student auch gleichzeitig Produzent.

3.2.5 Entitäten, Prozesse bzw. Aufgabenbereiche

Die *Entitäten* und die damit eingeschlossenen Prozesse in Langzeitarchivierungssystemen wie sie in Hochschul-Medienzentren Anwendung finden lassen sich entsprechend des OAIS-Referenzmodells klassifizieren in:

- Ingest (Einspeisen/ Aufnahme)
- Administration
- Datenmanagement (Verwaltung)
- Archivierung (Speicherung)
- Access (Verteilung)
- Preservation Planning (Bestandserhaltsplanung)

Feste Bestandteile der Langzeitarchivierungssysteme sind alle Entitäten bis auf Bestandserhaltsplanung. Dies ist in Ansätzen vorhanden, jedoch in derzeitigen Systemen nicht komplett integriert, was anhand der Beispielszenarien deutlich gemacht wird. Vielmehr ist hier ein Handlungsbedarf vorhanden. Die derzeitigen Systeme zur Langzeitarchivierung digitaler Information befinden sich in einer ununterbrochenen Weiterentwicklung und Umstrukturierung und es wird darauf bedacht, Bestandserhaltsplanung als eine funktionale Entität stärker zu integrieren, um einen Informationsverlust vorzubeugen. Gültige und allgemein anerkannte Konzepte oder Standardisierungen finden dafür derzeit keine Anwendung, jedoch aber individuelle Ansätze in den einzelnen Systemen.

3.2.6 Allgemeine Struktur Hochschul-Medienzentren

In Abbildung 14 ist die grundsätzliche Struktur der allgemeinen technischen Infrastruktur der Langzeitarchive von Hochschul-Medienzentren dargestellt. Dies dient der allgemeinen Charakterisierung der Systeme zur Langzeitarchivierung digitaler Information wie sie in Hochschul-Medienzentren Anwendung finden.

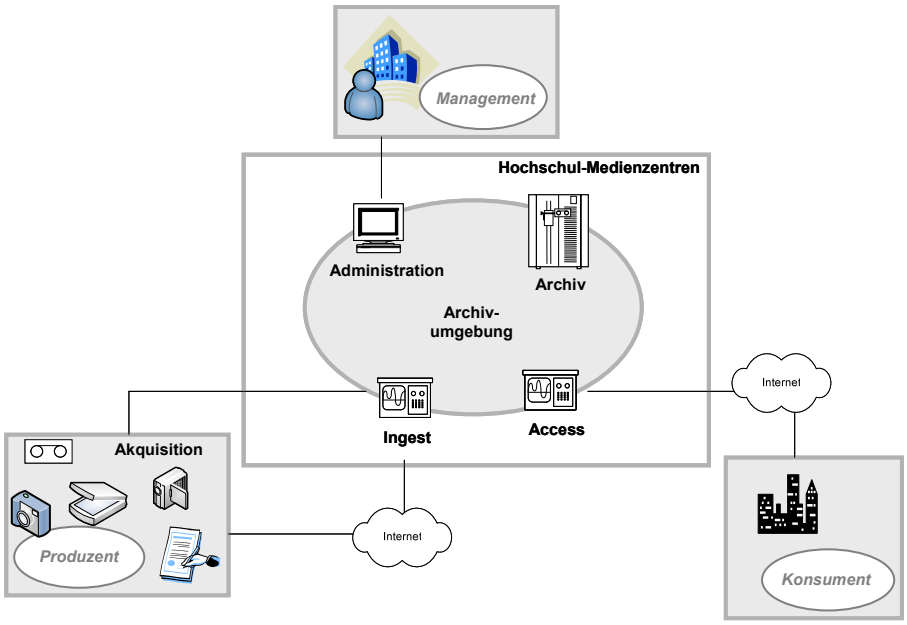

Abbildung 14: Grundsätzliche und allgemeine Struktur der allgemeinen technischen Infrastruktur der Langzeitarchive von Hochschul-Medienzentren.

Neue zu archivierenden Objekte werden akquiriert, d.h. gesammelt und aufbereitet. Sie werden dann in die Archivumgebung eingeliefert. Dies kann über eine direkte Verbindung geschehen oder über das Internet. In der Archivumgebung sind neben der Entität *Ingest* die Entität *Access*, die Entität *Administration* einschließlich *Data Management* und das eigentliche Archiv, die Entität *Archival Storage* vorhanden. Die allgemeine technische Infrastruktur ist oft so aufgebaut, dass die Prozesse, die mit *Ingest, Access* und *Data Management* verbunden sind von ein und derselben Entitäteninstanz, wie etwa ein Dokumentenserver ausgeführt werden.

3.2.7 Aufgaben

Das Management ist organisatorisch mit der Verantwortung sowohl für den ungestörten Ablauf der Prozesse innerhalb des Archivs als auch für die Regelung der Aufnahme von Archivpaketen sowie der Zugriffsregelung betraut. Die von Produzenten erstellte Information muss für die Archivierung aufbereitet werden, wobei in erster Linie Metadaten generiert werden und somit die Qualität der Archivpakete sichergestellt wird. Die aufgenommenen Archivpakete müssen dann verlustfrei und wieder findbar gespeichert und verwaltet werden. Darüber hinaus muss das Management den Lebenszyklus von Archivpaketen überwachen, wozu es Mechanismen des *Preservation Planning* anwenden soll, damit die Archivpakete auf lange Zeit verfügbar gemacht werden.

Systeme zur digitalen Langzeitarchivierung in Hochschul-Medienzentren sollten folgenden *Mindestsatz an Operationen* aufweisen:

- **Akquisition/ Sammlung**
  - Aufnahme von neuen originär in digitaler Form vorliegenden (*Born Digital*) Inhalten
  - Digitalisierung vorhandener analoger Information/ bereits vorhandener Archivalien (Digitalisate)
- **Dokumentation**
  - Katalogisierung der Archivalien. Der Katalog bestimmt wesentlich den Nutzwert eines Archivs.
- **Qualitätssicherung**
  - Generierung Metadaten
  - Aufbereitung der aufzunehmenden Archivalien für die Anforderungen des Archivs
- **Wiederverwertung – Access**

  - Auslieferung von Inhalten auf Anfragen
- **Management von Archivobjekten**
  - Verwaltung und Steuerung von Prozessen im Lebenszyklus von Archivobjekten
  - Zugriffskontrolle, Versionskontrolle und Rechtemanagement.

Diese Operationen beschreiben die grundsätzlichen Komponenten eines Archivsystems. Anforderungen der Vertrauenswürdigkeit und Sicherheit sind hier nicht explizit benannt, müssen aber in jeden einzelnen Punkt integriert werden.

3.2.8 Herausforderungen für die Systeme der Langzeitarchivierung im Einsatz in Hochschul-Medienzentren im Umgang mit verschiedenen Medien

Medien müssen in ihren unterschiedlichen Formaten erfasst und strukturiert verwaltet werden. Formate müssen mitsamt ihrer logischen Struktur mit Hilfe von technischen Metadaten im System hinterlegt sein. Konkrete Aspekte spielen dabei eine entscheidende Rolle:

- Identifizierung eines mehrfachen Auftretens eines digitalen Archivobjektes, Original, bitgenaue Kopie, Keyframe, komprimierte Vorschau, etc.
- Verwaltung der technischen Signaleigenschaften der Speicherung
- Verwaltung der strukturellen Eigenschaften der Speicherung
- Verwaltung von programmbeschreibenden Metadaten aus der Produktion
- Verwaltung der Metadaten (technische/ nichttechnische)
- Verwaltung von Rechtedaten
- Weitere Ergänzungen in den Datenbeständen für
  - einen verlustfreien Metadatenaustausch,
  - eine vollständige Metadatenerfassung
  - eine korrekte Darstellung der Archivobjekte und Inhalt
  - einen sicheren Zugriff auf Objekte und Metadaten

Ziele sind die Vermeidung von Medienbrüchen und die vollständige Erhaltung der Substanz auch über Transformationen hinweg, wie sie beispielsweise bei einer Migrierung nötig sind, hinweg. Ebenso müssen die Zusammenhänge von Multimediadaten hergestellt werden können. Auch dazu sind vollständige Metadaten notwendig, dies insbesondere bei der Erfassung und Aufnahme eines Informationsobjektes in das Archiv.

3.2.9 Allgemeine Archiveigenschaften

Während ein System zur digitalen Langzeitarchivierung in Rundfunkanstalten sehr komplex ist und aufgrund der ständigen Produktion dynamisch angelegt sein muss, ist ein System zur digitalen Langzeitarchivierung in Hochschul-Medienzentren weniger komplex und nicht derart hochdynamisch. Inhalte müssen nicht in Echtzeit abgefragt werden können und der Umfang der Produktion ist geringer. Zum Teil dient das eigentliche Archiv als *Endlager* für Archivobjekte, deren Abfragefrequenz gesunken ist und sie somit vom Cache-Zwischenspeicher (Dokumentenserver) genommen werden.

Generell wird der eigentliche Inhalt, die Essenz oder auch das digitale Archivobjekt genannt, getrennt von seiner Darstellung gespeichert. Die zur Darstellung notwendige Information wird in Metadaten mit und/ oder getrennt von dem Archivobjekt im Archiv gespeichert.

3.2.10 Medien und Daten in Langzeitarchiven der Hochschul-Medienzentren

Grundsätzlich wird in Archiven der Hochschul-Medienzentren Information der verschiedenen Medientypen digital gespeichert. Zu diesen Medientypen zählen Text, Bild, Audio, Video und vor allem Multimedia, d.h. es existiert eine Vielzahl von Darstellungsformen digitaler Information. Des

Weiteren ist Information zum einen digitalisiert und zum anderen von originär in digitaler Form vorliegender Natur, d.h. sie ist digital erstellt.

Im Zusammenhang mit der Langzeitarchivierung digitaler Information ist man aufgrund der verschiedenen Medientypen und der verschiedenen Darstellungsformen mit folgender Problematik konfrontiert: *Original und Kopie* bzw. Fälschung aufgrund von Manipulationen sind schwer zu unterscheiden.

Kopien sind auf der einen Seite gewünscht und unerlässlich für die abgesicherte und vertrauenswürdige Langzeitarchivierung. Auf der anderen Seite ist dadurch die Sicherung der Originalität in Frage gestellt. In digitalen Systemen sind Original und Kopie oft nur schwer, wenn überhaupt zu unterscheiden. Fälschungen und Manipulationen sind schwer nachvollziehbar. Die Frage ist nun: Braucht man wirklich ein Original für eine vertrauenswürdige Langzeitarchivierung und kann man dies in digitalen Systemen über einen langen zukünftigen Zeitraum sicherstellen? Für die Langzeitarchivierung digitaler Information müssen die folgenden Aspekte vom Archivmanagement berücksichtigt werden:

- Anzahl der Exemplare
- Kopien
- Original
- Auslieferung von Kopien
- Auslieferung orts- und zeitunabhängig
- Kontrolle der umlaufenden Kopien

3.2.11 Menge, Art und Ort anfallender Daten

Der Bestand in digitalen Langzeitarchiven in Hochschul-Medienzentren entstammt im Allgemeinen folgender *Herkunft*:

- Digitalisierung analogen Bestands
- Eigenerstellung
- Elektronische Publikationen
  - Kauf bzw. Lizenzierung (E-Books, E-Zeitschriften, usw.)
  - Freiwillige Abgabe/ gesetzlich geregelte Pflichtabgabe (amtliche Publikationen, Verlagspublikationen, usw.)
- Web-Ressourcen
- e-Learning Materialien

Der digitale Bestand setzt sich aus unterschiedlichen Medientypen zusammen. Digitale zu archivierende Materialien sind:

- **Selbsterstellte** Materialen, die in Bezug auf die künftige Verwendung nach entsprechenden Standards auf geeigneten Speichermedien mit präzisen Metadaten herstellbar und einsetzbar sind und
- **Fremderstellte** Materialien, die übergreifende Austauschformate brauchen, wobei offene und freie Standards eine entscheidende Rolle spielen und eine gemeinsame und möglichst standardisierte Metadatenstruktur von Vorteil ist.

Für den digitalen Bestand gilt, dass keine Medientypen ausgeschlossen werden bei der Bestandserschließung, Bestandsbereitstellung und Bestandserhaltung. Es ist jedoch angestrebt eine möglichst einheitliche und geringe *Anzahl von Medientypen* in einer Archivumgebung anzuwenden. Hochschul-Medienzentren sind weiterhin bestrebt digitale *Backupstrategien* bereitzustellen. Diese sind entweder lokal oder vernetzt und ermöglichen die Speicherung annähernd qualitativ gleichwertiger technischer Kopien. Generell soll auch ein Schutz der Integrität und Authentizität eines digitalen Objektes im Archiv gewährleistet sein, d.h. der unveränderte Erhalt des Sammelguts, der Schutz vor illegalen

Manipulationen, die Urheberschaft und Unversehrtheit sollen gesichert werden können. Da in digitalen Archivsystemen aufgrund einfacher Durchführbarkeit von spurlosen Veränderungen eine hohe Gefahr der Manipulierbarkeit vorliegt, muss das Archivmanagement, also die Hochschule angemessene Bestandssicherungsstrategien entwickeln und einsetzen. Solche Bestandstrategien müssen dabei gleichzeitig handhabbar und kostengünstig in der Durchsetzung sein.

Die anfallende Datenmenge hängt von der Art des Medientyps ab, also in welchem Format die Information gespeichert ist bzw. gespeichert werden soll.

Speichermedien und Speichermaterialien für multimediale Artefakte

Eingesetzte materielle *Trägermedien* liegen in verschiedenen Formen vor.

Anfallende Datenmenge

Der derzeitige Bestand von Hochschul-Medienzentren variiert und kann auch heute bereits bis zu mehrere PetaBytes an Materialien umfassen. Der Bestand in einem operativen Hochschul-Medienzentrum wächst ständig. Es kommen täglich neue originär in digitaler Form vorliegende Information (*Born Digital*) sowie neue Digitalisate hinzu, wobei alte, nicht mehr benötigte *Versionen* aus dem Archiv herausaltern. Weiterhin führen anfallende *Migrationen* zu einer erheblichen Erhöhung des Bestands. Die Bandbreite muss sich ebenfalls erweitern, da durch Audio und Video ein erhöhter Datenfluss und mit einer erhöhten Transferrate vorhanden ist. Die typische Bitrate für die Archivierung ist je nach Medientyp verschieden. Hochauflösendes Video ist hier nicht Schwerpunkt.

3.2.12 Eingesetzte technische Systeme und Art und Umfang der technischen Aufbereitung

Zunächst wird das Material durch den Produzenten gesammelt und aufbereitet. Die Einspielung des Materials in das Archiv (*Ingest*) erfolgt über eine Schnittstelle im Bereich der Produktion. Das Objekt bekommt bei Aufnahme eine *URN* zugewiesen, mit der es im Archiv wieder auffindbar sein wird. Das Objekt wird zu Archivierung weiter erschlossen, d.h. es werden Metadaten erzeugt und gespeichert. Im Archiv selbst besteht eine hierarchische Speicherstruktur, so dass das eingelieferte Objekt nicht automatisch sofort in das eigentliche Archiv, also den Archivspeicher wandert, sondern in einem Zwischenspeicher zwischengelagert wird.

3.2.13 Technische Infrastruktur

Die allgemeine technische Infrastruktur einer Umgebung zur Langzeitarchivierung digitaler Multimediainhalte ist in Abbildung 15 dargestellt und lässt sich aufteilen in die Bereiche: Sammlung und Erstellung, Ingest einschließlich Erschließung, Archiv und Access mit Zugriffsportal. Innerhalb der Bereiche kommen auf die jeweiligen Aufgaben spezialisierte Hard- und Software zum Einsatz. Solche Systemkomponenten als auch die Bereiche, denen sie angegliedert sind, sind über Informationstechnik-Netzwerke miteinander verbunden.

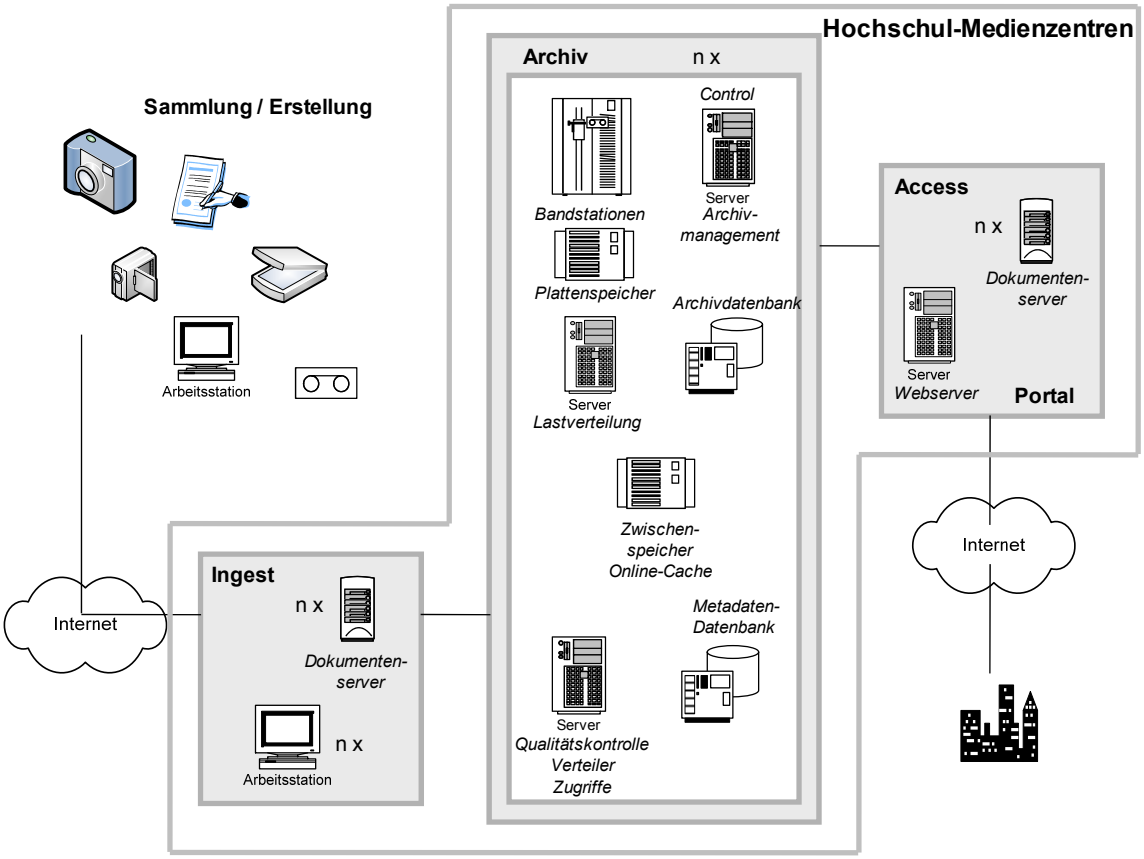

Abbildung 15: Allgemeine technische Infrastruktur mit grundsätzlichen Komponenten der Hochschul-Medienzentren.

Systemkomponenten

Typischerweise kommen folgende technische Hardwaresysteme in der Archivumgebung von Hochschul-Medienzentren zum Einsatz.

Hardware:

- Bandbibliotheken (*Tape Libraries*), Bandlaufwerke und Bandroboter
- Plattenspeichersysteme
- Verteilte Medienspeicher
- Dokumentenserver
- Kontrollserver
- Server zur Lastenverteilung
- Anwendungsserver
- Netzwerk
- Arbeitsplatzrechner
- Digitalisierungsautomationen

3.2.14 Technische Plattformen, deren Eigenschaften und Mechanismen zur Speicherung von hochkomplexen und verteilt liegenden Inhalten

Einzelne Speichersysteme

Eine verteilte Speicherung betrifft in Hochschul-Medienzentren bisher vorrangig die einzelnen Speichersysteme, die eine *Clusterarchitektur* aufweisen und aus einer bestimmten Anzahl von Knoten (*Cluster*) mit Speicher-, Rechen- und Kommunikationskapazität bestehen. Hier kommt es auf die Clusterverwaltung an, um ihre Leistung maximal ausnutzen zu können. Zur Clusterverwaltung zählen eine effiziente Lastenverteilung, die Auftragspriorisierung und Ressourcenplanung. Knotenausfälle werden über regelmäßige Kontrollsignale automatisch erkannt und die Aufgaben des ausgefallenen Knoten werden auf andere Knoten verteilt.

Verschiedene Datenbänder

Inhalte werden in Hochschul-Medienzentren bereits zum Teil auf verteilte Datenbänder gespeichert. Dies jedoch dann, wenn sie sich im ständigen Zugriff befinden, also auf Dokumentenserver, da hier ein erhöhter Durchsatz anliegt. In einer Datenbank sind die Speicherorte zur Wiederauffindbarkeit vermerkt.

Geographische Standorte

Hochschul-Medienzentren betreiben in der Regel einen *Hauptstandort* für ihr digitales Langzeitarchiv. Darüber hinaus haben sie, insofern die Kosten es zulassen, ein oder mehrere *Backups* des Archivs an verschiedenen geographischen Orten. Das Archiv des Hauptstandortes muss mit den Backups synchronisiert werden. Dies wird über so genannte *Control-Server* verwaltet und gesteuert.

Software

- Archivmanagement
- Dateitransfer
- Transcodierer
- Erschließungssoftware – Qualitätswerkzeuge
- Clients
- Webportale
- Datenbanken
    - Archivdatenbanken
    - Metadaten-Datenbanken einschließlich Rechtedaten

Qualitätswerkzeuge dienen der Erschließung von eingelieferten Archivobjekten. An dieser Stelle bleibt offen, welche Qualitätswerkzeuge das sind. Zum einen sind es Werkzeuge zur Generierung von Metadaten, um ein Objekt im Archiv wieder auffindbar zu machen. Zum anderen müssen es Werkzeuge der IT-Sicherheit sein, die bereits bei der Erschließung die Integrität und Authentizität sichern.

3.2.15 Inhaltsbeschreibungen und Rechtedaten sowie Metadaten

Metadaten sind entscheidend für die Verfügbarkeit von digitalen Archivobjekten über eine lange Zeit sowie für deren Vertrauenswürdigkeit und damit für eine Langzeitarchivierung digitaler Information unabdingbar. Metadaten enthalten Information über das digitale Archivobjekt. Metadaten sollen sicherstellen, dass Information langfristig interpretiert und genutzt werden kann. Metadaten werden unterschieden in technische Metadaten, Metadaten zu Dateien, strukturelle Metadaten, administrative Metadaten und rechtliche Metadaten. Es handelt sich hierbei durchgängig um objektbezogene Metadaten:

Technische Metadaten beschreiben die erforderliche technische Umgebung u.a. das Darstellungsprogramm, welches notwendig ist, ein Archivobjekt zu interpretieren und lesbar darzustellen. Für die automatische Generierung von derartigen technischen Metadaten ist das Werkzeug JHOVE

(JSTOR/Harvard Object Validation Environment)[9] entwickelt worden und in Systemen zu Langzeitarchivierung in Hochschul-Medienzentren im Einsatz. Mittels dieses Werkzeugs wird das vorliegende Format beim Einspeisen ins Archiv identifiziert, validiert und charakterisiert und es werden technische Metadaten erzeugt und abgelegt.

Metadaten zu Dateien enthalten eine oder mehrere Elemente, um zusammengehörige Dateien, die getrennt im Archiv gespeichert sind, zusammenzuhalten. Ein Element listet alle Dateien auf, die eine einzelne komplette elektronische Version eines digitalen Objekts ausmachen. Dies ist gerade in hochgradig verteilten Speichersystemen notwendig.

Strukturelle Metadaten beschreiben Relationen zu anderen Objekten und Ressourcen im Archiv. Strukturelle Metadaten beschreiben aber auch die hierarchische Struktur, die den Nutzern eines digitalen Objektes zur Navigation innerhalb des Objektes angeboten werden kann [LoC05].

Administrative Metadaten beschreiben den Lebenszyklus des Objekts und bestimmen darüber hinaus auch Archivierungsmaßnahmen. Administrative Metadaten beinhalten Information zur Verwaltung des ursprünglichen Archivobjektes. Diese werden unterteilt in [LoC05]:

- Technische Metadaten – Angaben über die Herstellung der Dateien, das Format und Nutzungsbesonderheiten
- Urheberrechte und Lizenzinformation (Nutzungsrechte)
- Angaben zur Quelle – Beschreibungs- und Verwaltungsangaben zur analogen Quelle einer Digitalisierung
- Digitale Herkunftsangaben – zum Quelle- und Abbild-Verhältnis zwischen Dateien, einschließlich der Bestimmung von Master und Derivaten und Informationen über Migrationen und Transformationen der Daten zwischen der ursprünglichen Digitalisierung eines Gegenstandes und seiner gegenwärtigen Erscheinung als digitales Objekt

Rechtliche Metadaten beschreiben Zugriffsmodalitäten auf ein digitales Archivobjekt einschließlich Urheberrechte und Nutzungsrechte.

Beschreibende Metadaten dienen der Erschließung langfristigen Auffindbarkeit von Ressourcen und von Archivobjekten.

Persistent Identifier

Persistente Identifier (PI) sind eindeutige, standortunabhängige Identifikatoren für digitale Objekte. Sie sind ein wichtiger Baustein für die langfristig stabile Adressierung von digitalen Objekten. Generell wird damit eine Trennung von Standortreferenz und Identifikation bezweckt. Die Ressource selbst bzw. das Archivobjekt sollen so identifiziert werden, nicht der Standort, der sich in einem Langzeitarchiv immer wieder ändern kann. Ein PI wird über Resolving-Dienst in eine Standortreferenz (URL) aufgelöst. Ändert sich Standort muss nur Eintrag im Resolver geändert werden, PI bleibt gleich. Sie ist somit global im Archiv gültig. Persistente Identifikatoren dienen der eindeutigen Identifikation und Authentifikation von Ressourcen und Archivobjekten und unterstützen das Einbinden von Rechten und Zugriffsstrukturen. Sie werden bei der Erschließung des zu archivierenden Objektes vergeben und bleiben unverändert für dieses Archivobjekt bestehen.

Spezifische Anforderungen an persistente Identifikatoren sind:

- Eindeutigkeit
- Dauerhaftigkeit
- Globalität
- Integration in Benutzungsschnittstelle
- Transferprotokoll
- Zentrale Registrierung
- Identifikatorenverzeichnis

[9] http://hul.harvard.edu/jhove

DOI[10] – Digital Object Identifier – ist ein System zur Identifizierung von Inhaltsobjekten in einer digitalen Umgebung. DOI® Namen sind verbunden mit allen Entitäten, die in digitalen Netzwerken fungieren, und ändern sich nicht, auch wenn sich die digitale Entität ändert. Weitere Ausführungen sind unter der angegebenen Referenz zu finden.

Weitere Information zu persistenten Identifikatoren ist beispielsweise auf der Website Persistent Identifier[11] zu finden.

Format Registries

Format Registries im Zusammenhang der Langzeitarchivierung sind Datenbanken und enthalten umfassende und in standardisiert erfasster und auslesbarer Form Information zu Dateiformaten. Diese Information enthält u.a. Namen, Versionierung, Zeichensatz und Hinweise zu Hard- und Softwareanforderungen. Hier sind z.B. PRONOM[12] oder das Global Digital Format Registry (GD-FR)[13] zu nennen.

Metadaten-Spezifikationen

Bezüglich der Metadaten-Definitionen gibt es verschiedene, miteinander kooperierende Organisationen und Standardisierungen, wovon hier eine Auswahl genannt ist.

OCLC/RLG[14] – *Preservation Metadata Working Group* ist fortgeführt durch **PREMIS** (*PREservation Metadata Implementation Strategies*) [PREMIS05]. Ziel dieses Rahmenkonzeptes für Metadaten zur Langzeiterhaltung ist die Schaffung von implementierbaren Langzeitarchivierungsmetadaten mit größtmöglicher Anwendbarkeitsbreite.

Metadata Encoding and Transmission Standard (METS)[15] [LoC05] – METS ist ein von der Digital Library Federation (DLF) geförderter XML-basierter Standard zur Speicherung von digitalen Objekten mit ihren Meta- und Strukturdaten. METS ist ein Containerformat für digitale Objekte. Unterschiedliche Strukturen (logische, physische etc.) können abgebildet und unterschiedliche Metadatenstandards können in METS berücksichtigen werden. METS bietet demnach eine hohe Flexibilität und eignet sich dafür, einheitliche Container als *Submission Information Package (SIP)* im Sinne des OAIS zu definieren. [Alt05]

Metadata Standards Framework[16] – Die derzeit konkreteste und umfassendste Beschreibung von Metadaten zur Langzeitarchivierung ist im *Metadata Standards Framework* der National Library of New Zealand festgehalten. Ende 2002 folgte dem Rahmenwerk eine konkrete Implementierung von Preservation Metadata.

Langzeitarchivierungsmetadaten für elektronische Ressourcen (LMER)[17] ist eine Entwicklung der Deutschen Bibliothek. Es wird ein Kern von relevanten technischen Metadaten zur Langzeitarchivierung definiert. LMER ist XML-basiert und als Austauschformat anzusehen. Es ist universell ausgerichtet.

Dublin Core (DC)[18] – Der *Dublin Core* Metadatenstandard ist ein etablierter internationaler Standard (ISO 15836) zur Beschreibung von digitalen Objekten. Weitere Information ist unter der angegebenen Referenz zu finden.

---

[10] http://www.doi.org/

[11] http://www.persistent-identifier.de/

[12] http://www.nationalarchives.gov.uk/PRONOM/default.htm

[13] http://hul.harvard.edu/gdfr/

[14] http://www.oclc.org

[15] http://www.loc.gov/standards/mets

[16] http://www.natlib.govt.nz/files/4initiatives_metaschema.pdf

[17] http://www.d-nb.de/standards/lmer/lmer.htm

[18] http://dublincore.org/

OWL[19] – Der *Standard Web Ontology Language (OWL)* dient im Zusammenhang mit Webservices der formalen Beschreibung von Ontologien, so dass Software deren Bedeutung verarbeiten kann.

MPEG-7[20]**/ MPEG-21**[21] – Die Standards MPEG-7 und MPEG-21 sind von der *Moving Picture Experts Group* in Bezug auf Multimedia spezifiziert. MPEG-7 (*Multimedia Content Description Interface Standard*) dient dazu, Information über den Inhalt bereitzustellen. Um die audiovisuellen Inhalte unabhängig von ihrer Speicherung, Darstellung, und Übertragung sowie unabhängig vom Medium oder den Technologien zu beschreiben, beinhaltet MPEG-7 einen umfassenden Satz von standardisierten Werkzeugen[22]. Auf diese Weise ist eine einfache, flexible und interoperable Lösung geschaffen für die Problematik des Erschließens und Katalogisierens (*Indexing*), Suchens und Abfragens (*Retrieval*) von Multimedia. Mit dem verwandten ISO/IEC-Standard MPEG-21 ist ein Multimedia-Rahmenwerk definiert, um die Prozesse zur Bereitstellung von Inhalten zu unterstützen. Der Standard MPEG-21 wurde spezifiziert, um den Austausch, den Zugriff und Handel, das Konsumieren und das Verändern von digitalen Multimediaangeboten auf effizientem, transparentem und interoperablem Weg zu ermöglichen.

RDF[23] – *Resource Description Framework* ist eine Infrastruktur, die Metadaten im Stil des *Dublin Core* bereitstellt. Die Infrastruktur schließt eine Darstellung von Site Maps und Taxonomien ein. Es ist eine XML-basierte Sprache zur Spezifikation von Graphen (vgl. semantischer Netzwerke). Weitere Ausführungen sind unter der angegebenen Referenz zu finden.

Speicherung und Bereitstellung von Metadaten

Metadaten sind zum einen zusammen mit dem Archivobjekt gespeichert und zum anderen getrennt vom Archivobjekt.

[19] http://www.w3.org/2007/OWL/wiki/OWL_Working_Group

[20] http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

[21] http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm

[22] http://xml.coverpages.org/mpeg7.html

[23] http://www.w3.org/RDF/

4 Langzeitarchivierungstechniken in den Szenarien - Systemabstraktion mit Zuordnung der Anforderungen und Annahmen

In diesem Kapitel erfolgt für jedes Szenario die Abstraktion der exemplarischen Langzeitarchivierungs-Einzelsysteme auf Basis des OAIS-Referenzmodells. Die geleistete Abstraktion unterscheidet Akteure im Umfeld des Archivs, Architektur und Rollen sowie Daten. Die Darstellung der Informationsflüsse betrachtet die drei bedeutenden Vorgänge **Bestandserweiterung**, **Bestandsbereitstellung** und **Bestandserhaltung**. Diese Vorgänge umfassen zusammenhängende Abläufe zwischen den Akteuren Produzent, Management, Konsument und Archiv über Systemgrenzen hinweg und unter Einbezug der funktionalen Entitäten des Archivs. Darauf aufbauend erfolgt eine Zuordnung der Anforderungen und Annahmen für die exemplarischen Szenarien als vertrauenswürdige und abgesicherte Langzeitarchivierungssysteme.

4.1 Öffentlich-rechtliche Rundfunkanstalten

4.1.1 Akteure

Aufgrund ihrer engen Einbindung in nahezu alle Arbeitsschritte des Fernsehproduktionsprozesses ist die Anzahl der mittelbaren menschlichen Nutzer des Archivs eine Vielzahl (Abbildung 16). Sie lassen sich – ohne Anspruch auf Vollständigkeit – in Nutzergruppen gemäß der Organisation der Rundfunkanstalt gliedern in Redaktion, Produktion, Postproduktion, Distribution und Archiv. Einzelne Akteure sind Redakteur, Cutter, Sendeautomation, Archivar, Dokumentar und Weitere. Oftmals greifen diese Akteure sowohl schreibend als auch lesend auf das Archiv zu, nehmen also beide Rollen Produzent und Konsument ein. Archivare der Rundfunkanstalt, aber auch technische Administratoren, sind als Akteur in der Rolle des Managements einzustufen. Sämtliche Akteure sind Angehörige der Rundfunkanstalt.

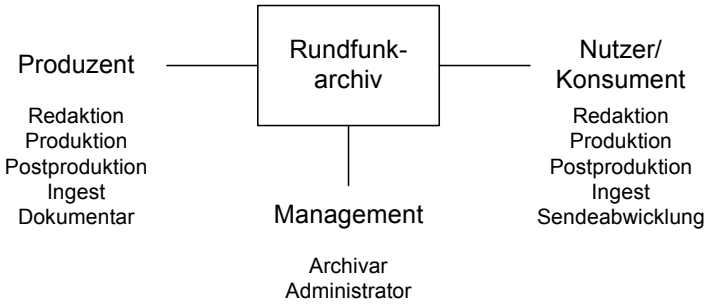

Abbildung 16: Akteure im Umfeld des Rundfunkarchivs.

4.1.2 Architektur und Rollen

Anhand der Organisation der Rundfunkanstalten lassen sich die Bereiche Redaktion, Produktion-&-Postproduktion, Ingest, Portal und Sendeabwicklung als Akteure im Umfeld des Archivs identifizieren (Abbildung 17). Die technischen Systeme dieser Bereiche interagieren mit den technischen Systemen des Archivs.

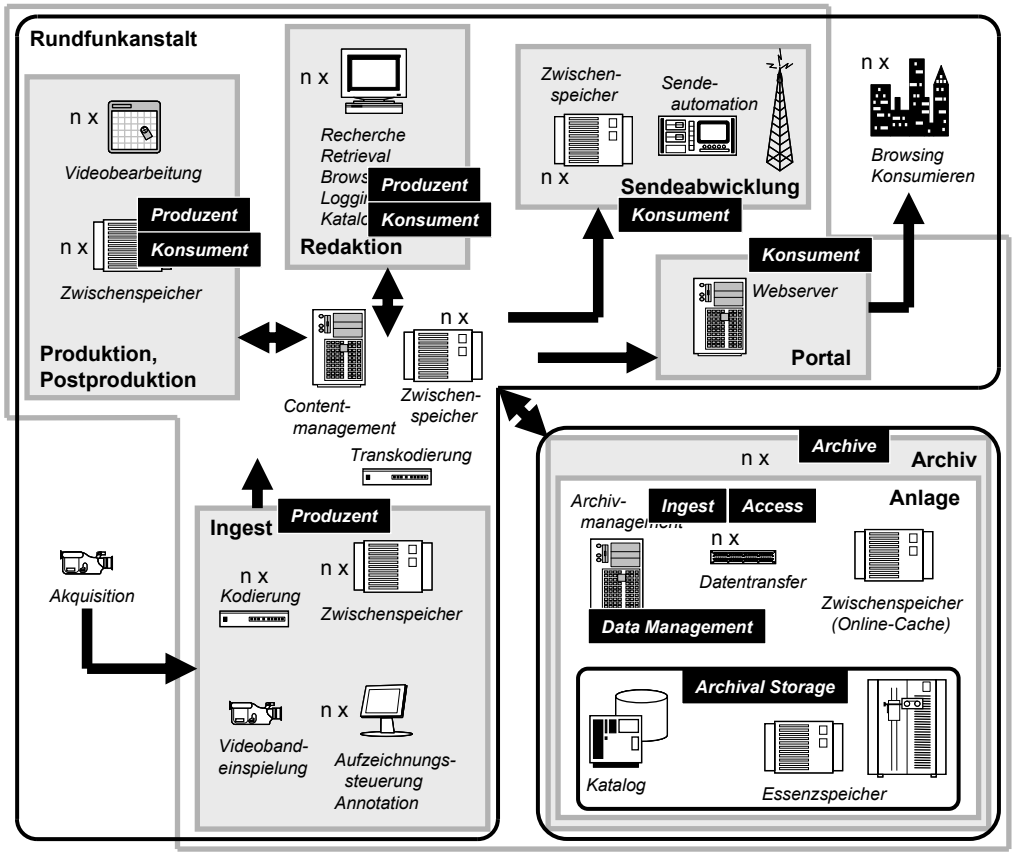

Abbildung 17: Auf das Archiv bezogene Rollen und Informationsflüsse in einer vernetzten Produktionsumgebung in den Rundfunkanstalten.

Die Bereiche stehen dem Archiv jedoch nicht einzeln in der Rolle als Produzent respektive Konsument gegenüber. Ursächlich ist die Vermaschung der vernetzten Produktionsumgebung mit ihrer gemeinsamen Verwaltung verteilt und mehrfach vorliegender Inhalte einschließlich der Steuerung und Automatisierung von Arbeitsabläufen. Das Content-Management als zentrale Komponente vermittelt. Ein weiterer wesentlicher Faktor ist die logische Unterscheidung zwischen Essenz und Metadaten. Während ein so genanntes Datenmanagement primär die formalen und beschreibenden Metadaten behandelt, verwaltet und steuert das Essenzmanagement sämtliche zur Essenz gehörenden Medien. Aufgrund der deutlich unterschiedlichen Eigenschaften sind auch operativ verschiedene Systemkomponenten für deren Handhabung zuständig. Während regulären Anwendungsservern Verwaltungs- und Steuerungsaufgaben anhand der Metadaten zukommen, sorgen Verteilte-Festplattenspeichersysteme und Medienserver respektive Videoserver für Lagerung und Verlagerung der Essenzen. Erschwerend kommt hinzu, dass ein Material anfragender Konsument nicht zwangsläufig identisch ist mit dem Empfänger des Archivmaterials. In Folge dessen kann die gesamte Produktionsumgebung der Rundfunkanstalt als ein umfassender *Produzent und Konsument* gegenüber dem Archiv aufgefasst werden.

Innerhalb des Archivs entspricht die technische Infrastruktur im Wesentlichen dem Referenzmodell. Bandstationen und Plattenspeicher sind dem *Archival Storage* zuzuordnen. Ihnen vorgeschaltet ist oftmals ein Zwischenspeicher. Das Archiv-Management auf einem Anwendungsserver, unterstützt durch Katalogdaten auf einem Datenbanksystem, regelt das *Data Management*. Die Unterscheidung nach Metadaten und Essenz hat zur Folge, dass jeweils mehrere Systeme *Ingest* und *Access* be-

wältigen. Das Archiv-Management handhabt die Verwaltung von Bestandsbereitstellung und Bestandserweiterung anhand der Metadaten und steuert Datenpumpen (Aktoren) mit der Einspielung bzw. Ausspielung der umfangreichen Essenzdaten.

Die Vorgänge der *Einlieferung (Submission)* und *Auslieferung (Dissemination)* von Material in das Archiv bzw. aus dem Archiv können im Rahmen der zentralen Speicherverwaltung als eine Verlagerung von Material zwischen Online-/ Nearline- und der Archivspeicherebene aufgefasst werden. Dem oftmals vorhandenen Zwischenspeicher der Archive kommt die Rolle der Nearline-Ebene zu.

4.1.3 Informationsflüsse

Die Betrachtung der Informationsflüsse mit Bezug auf das Archiv basiert auf einer abstrahierten technologischen Architektur der Rundfunkanstalten. Betrachtungsgegenstand sind Bestandserweiterung, Bestandsbereitstellung sowie Bestandserhaltung. Weil die Archive der Rundfunkanstalten eng in die vernetzte Produktionsumgebung eingebunden sind und Havarielösungen zusätzliche alternative Arbeitsabläufe vorsehen, ist die Menge der Informationsflüsse mit Bezug auf das Archiv Vielzahl. Die in dieser Expertise nachfolgend behandelten Vorgänge stellen eine exemplarische Auswahl aus dem Regelbetrieb in den Rundfunkanstalten dar.

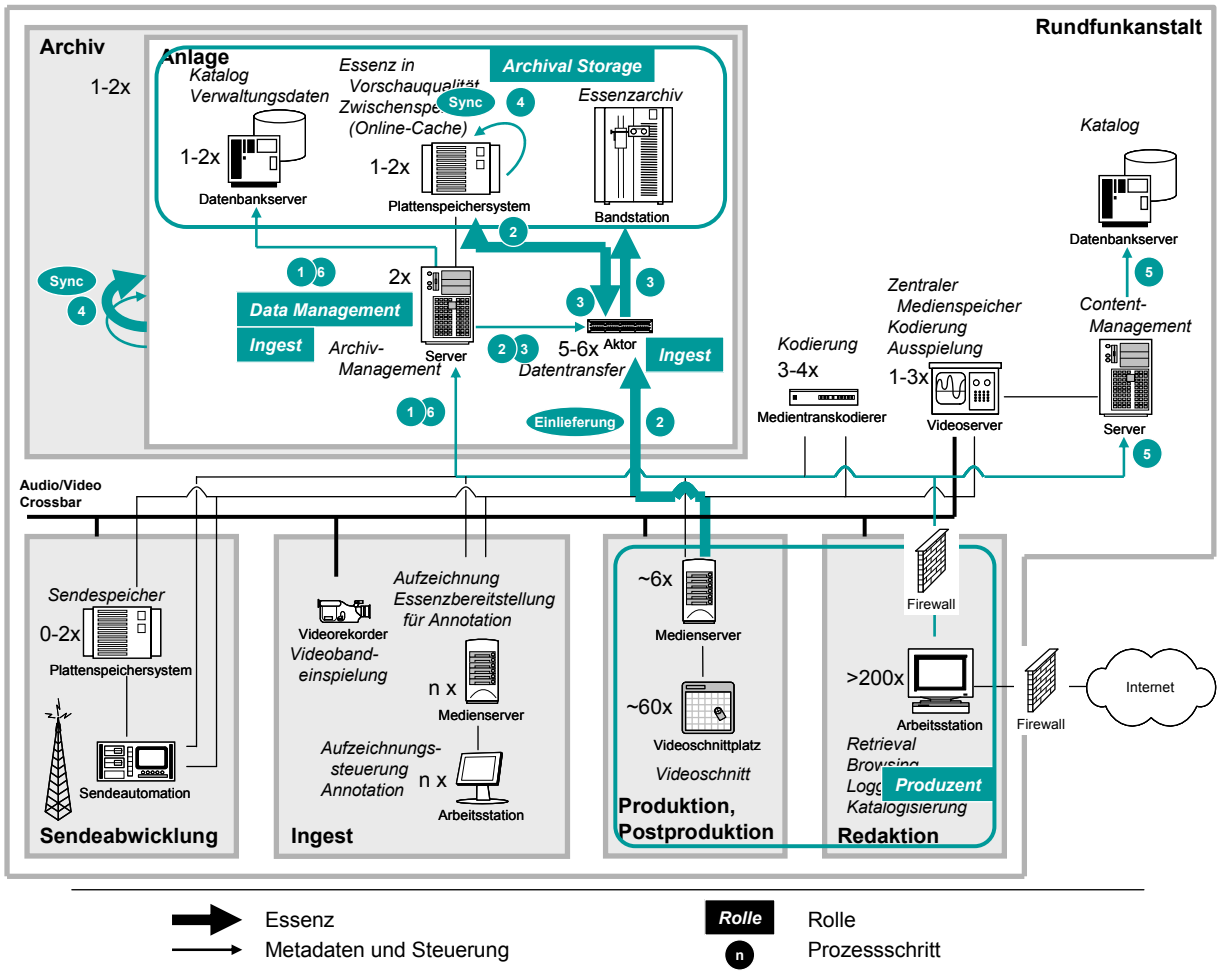

Abbildung 18: Rollen und Informationsflüsse bei einer Bestandserweiterung im abstrahierten System des Beispielszenarios öffentlich-rechtliche Rundfunkanstalten.

Bestandserweiterung

Beispielhaft für die Einlieferung (*Submission*) von Material in das Archiv wird nachfolgend die Archivierung der Endfassung eines Fernsehbeitrages aus der Postproduktion behandelt (Abbildung

18). Weitere übliche Einlieferungsfälle stellten die Aufzeichnung des Sendestroms aus der Sendeabwicklung oder die Rückwärtsdigitalisierung von analogen Archivbeständen durch den Ingest dar.

Tabelle 4: Prozessablauf bei einer Bestandserweiterung im abstrahierten System des Beispielszenarios öffentlich-rechtliche Rundfunkanstalten.

Prozessschritt Nr.	Vorgang
1.	Auslösen des Archivierungsvorgangs durch das Content-Management. Übertragung der Metadaten und des gegenwärtigen Speicherortes der zu archivierenden Essenz.
2.	Kopieren von Material von den Medienservern der Postproduktion zum Zwischenspeicher des Archivs. Hierzu Transferauftrag an eine Datenpumpe (Aktor).
3.	Unmittelbar anschließend wird über vordefinierte Speicherregeln, gesteuert vom Archivmanagement, das Material sofort in die Robotik des Archivs geschrieben. Hierzu Transferauftrag an eine Datenpumpe (Aktor).
4.	Sofern vorhanden wird ausgehend vom Zwischenspeicher-Inhalt eine Kopie in einer zweiten Archivanlage ausgelöst. Erstellung einer kompletten Spiegelung neuer Archivalien durch das Archivmanagement. Sofern vorhanden, Spiegelung der Vorschauqualität auf einem zweiten Festplattenspeichersystem.
5.	Inhaltliche Erschließung über eine Arbeitsstation. Einpflegen der Metadaten in das Content-Management.
6.	Aktualisierung der Katalog-Metadaten des Archivs mittels Änderungsbenachrichtigung vom Content-Management.

Bestandsbereitstellung

Beispielhaft für die Bestandsbereitstellung von Material aus dem Archiv wird nachfolgend die manuelle Recherche und anschließende Bereitstellung von Archivmaterial für die Produktion behandelt (Abbildung 19). Einen weiteren üblichen Auslieferungsfall stellte die automatisierte Anforderung von Material anhand von Transmissionslisten für den Sendebetrieb dar.

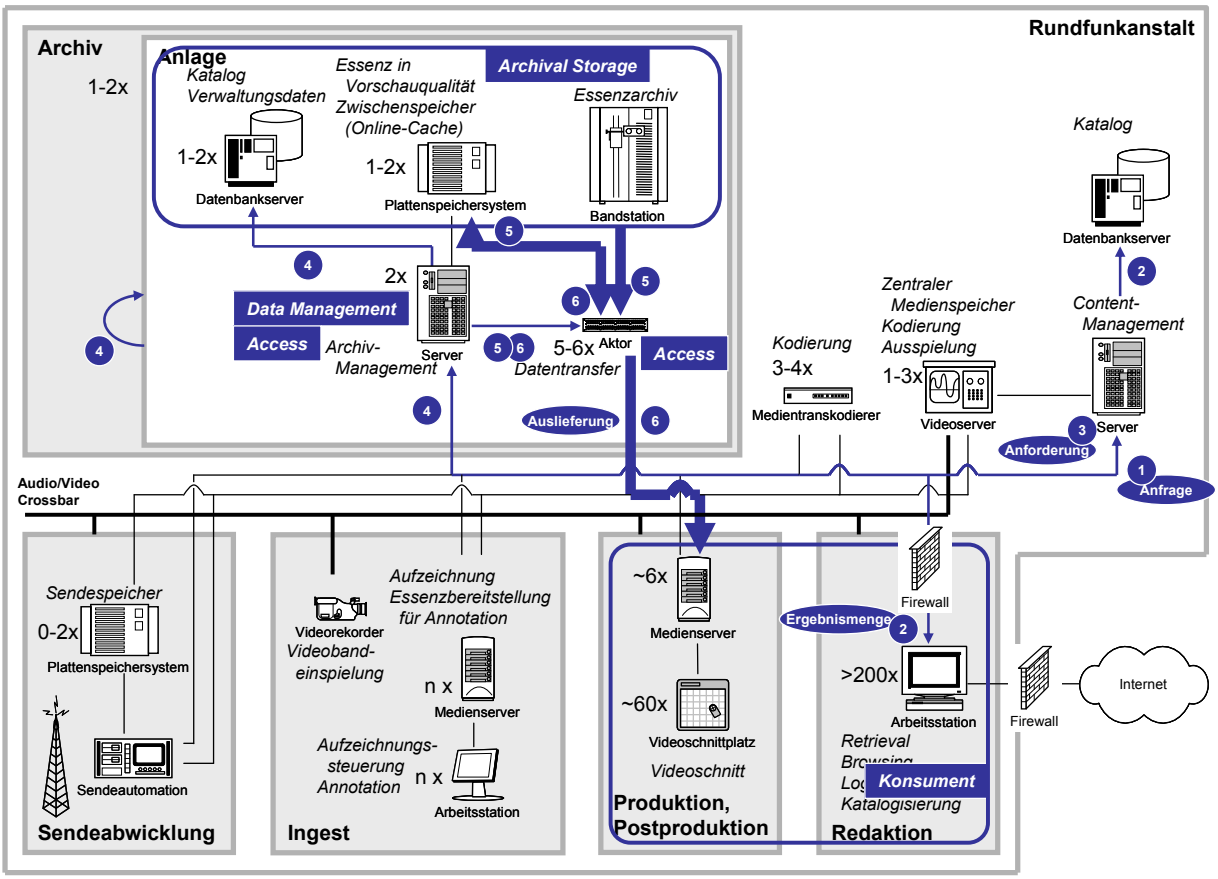

Abbildung 19: Rollen und Informationsflüsse bei einer Bestandsbereitstellung im abstrahierten System des Beispielszenarios öffentlich-rechtliche Rundfunkanstalten.

Tabelle 5: Prozessablauf bei einer Bestandsbereitstellung im abstrahierten System des Beispielszenarios öffentlich-rechtliche Rundfunkanstalten.

Prozessschritt Nr.	Vorgang
1.	Recherche im Materialbestand der vernetzten Produktionsumgebung über eine Arbeitsstation. Stellen von Anfragen an das Content-Management.
2.	Ermittlung zutreffender Katalogeinträge und Übermittlung der Ergebnismenge an die Arbeitsstation.
3.	Auswahl und Anforderung von Essenz über das Content-Management.
4.	Sofern das Material nicht bereits auf dem zentralen Speicher oder Medienservern der Produktion vorliegt, Weiterleitung als Materialanfrage an das Archiv-Management. Ein Lastverteilmechanismus erlaubt die gleichzeitige Wiederherstellung aus gespiegelten Archivanlagen.
5.	Sofern das Material nicht bereits im Archiv-Zwischenspeicher vorliegt, Bergung der Essenz von Datenband und Überspielung in den Zwischenspeicher des Archivs. Hierzu Transferauftrag an eine Datenpumpe (Aktor).
6.	Fristgerechte Auslieferung zum Bereitstellungstermin an das ausgewählte Zielsystem. Hierzu Transferauftrag an eine Datenpumpe (Aktor).

Bestandserhaltung

Archive der Rundfunkanstalten treffen eine Reihe von Maßnahmen zur Erhaltung ihrer Bestände. Im Folgenden seien damit verbundene Informationsflüsse wiedergegeben (Abbildung 20).

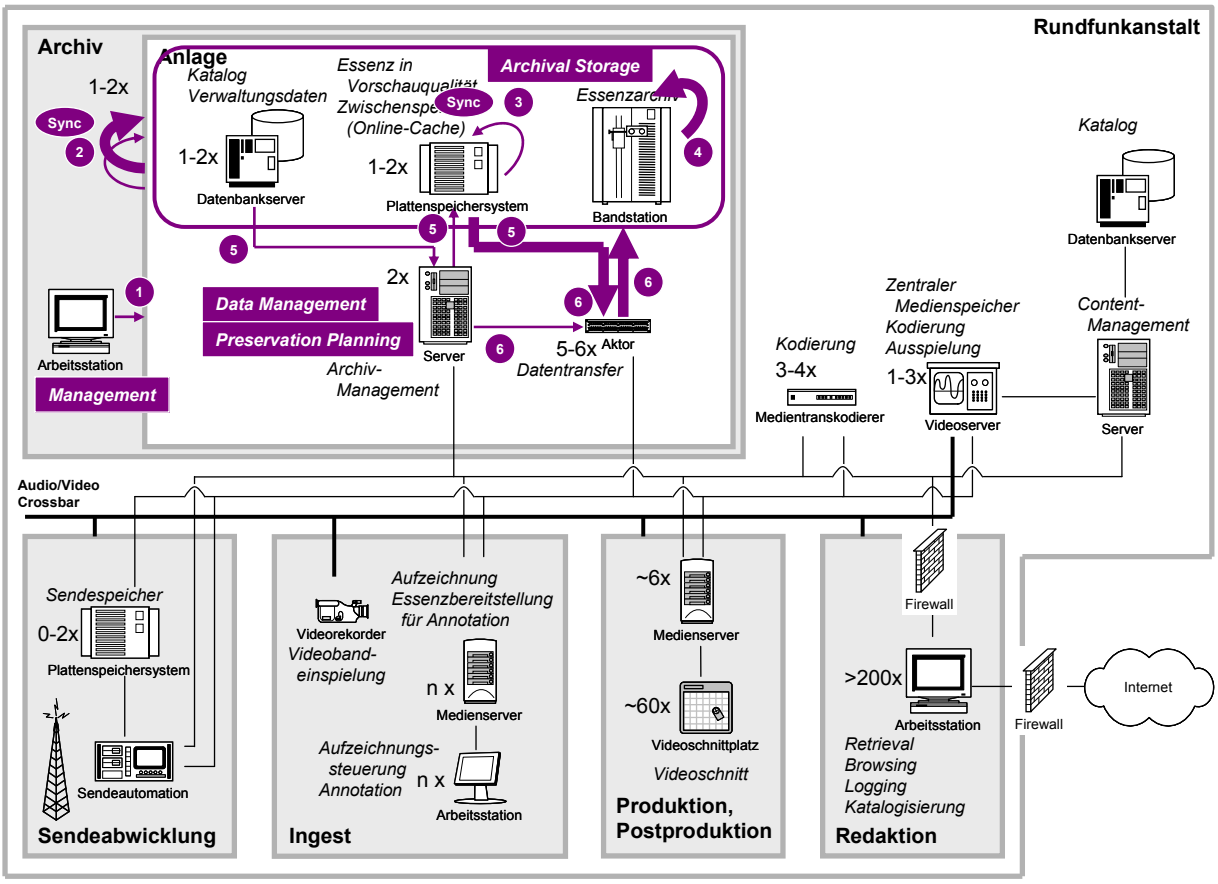

Abbildung 20: Rollen und Informationsflüsse bei der Bestandserhaltung im abstrahierten System des Beispielszenarios öffentlich-rechtliche Rundfunkanstalten.

Tabelle 6: Maßnahmen bei der Bestandserhaltung im abstrahierten System des Beispielszenarios öffentlich-rechtliche Rundfunkanstalten.

Maßnahme Nr.	Vorgang
1.	Vorgabe der Maßnahmen zur Bestandserhaltung durch das Management.
2.	Spiegelung der Archivinhalte und Katalogdaten in einer zweiten Anlage.
3.	Spiegelung der Essenz in Vorschauqualität und Schlüsselbilder (*Keyframe*) auf zweitem Plattenspeichersystem.
4.	Umkopieren der Inhalte von als (potenziell) schadhaft erkannten Datenträgern.
5.	Datensicherung (*Backup*) der Katalogdaten auf Bandrobotik.
6.	Datensicherung (*Backup*) der Essenz in Vorschauqualität und Schlüsselbilder (*Keyframe*) auf Bandrobotik.

4.1.4 Daten

Video- und Audiomaterial in Produktionsqualität (*HighRes*), kurz Essenz genannt, ist in den Rundfunkarchiven die aus Sicht des Referenzmodells zu bewahrende *Content Information*. Die für die korrekte Interpretation notwendige *Representation Information* besteht in erster Linie aus Audio- und Videoformatangaben. Die Essenz in Vorschauqualität (*LowRes*) dient ebenso wie die Schlüsselbilder (*Keyframes*) dem Auffinden von Archivbeständen und ist wie die Metadaten aus der inhaltlichen Erschließung als *Descriptive Information* einzuordnen.

Eine Zusammenführung von Content Information und *Preservation Description Information (PDI)* zu einem *Information Package* findet im Rundfunkbereich ausschließlich auf konzeptueller Ebene statt. Gleiches gilt für die Verknüpfung von Descriptive Information mit den von ihr beschriebenen Inhalten. In den untersuchten Systemen ist die getrennte Behandlung von Essenz – gleich Content Information – einerseits und Metadaten – entspricht Preservation Description Information (PDI), Representation Information sowie Descriptive Information – andererseits stark ausgeprägt.

Grundsätzlich lässt sich auch im Rundfunkbereich eine Unterscheidung nach *Submission Information Package (SIP)*, *Archival Information Package (AIP)* und *Dissemination Information Package (DIP)* nachvollziehen. So gilt es beispielsweise, aufgrund der Heterogenität der Teilsysteme in der vernetzten Produktionsumgebung, Metadaten zwischen verschiedenartigen Schemata zu übersetzen. Auch ist gegebenenfalls das Repräsentationsformat des Content für eine Auslieferung abzuändern. Und der Metadatensatz einer Essenz in der Produktion ist oftmals um ein Vielfaches umfangreicher als jener von Archivalien. Für sämtliche Information Packages gilt, dass Content und Metadaten getrennt verarbeitet werden.

4.1.5 Annahmen über die vertrauenswürdige und abgesicherte Langzeitarchivierung

Für die weitere Auswertungsarbeit trifft diese Expertise nachfolgende *Annahmen* über die Beispielsysteme der öffentlich-rechtlichen Rundfunkarchive.

In der vernetzten Produktionsumgebung ist eine zentrale Instanz für die Verwaltung der Inhalte einschließlich Metadaten zuständig und dient als Verteiler (Abbildung 10). Diese Expertise setzt das Archiv *nicht* dieser zentralen Position gleich. Sie betrachtet im Folgenden das Archiv als eine den anderen Bereichen der Rundfunkanstalt gleichberechtigte funktionale Einheit, welcher die dauerhafte Verwahrung wesentlicher Produktionsergebnisse und deren Bereitstellung für Folgeproduktionen zukommt. Dieser Art ist das Archiv in die Produktionsprozesse eingebunden, ohne dass jedoch jede Aktualisierung gemäß der Philosophie der ständigen Informationsanreicherung entlang der Produktionskette von ihm zu behandeln wäre. Dies schließt eine Änderung und Aktualisierung von Archivalien nicht aus.

Die Darstellungen von Architektur und Informationsflüssen in den Rundfunkanstalten stellen eine Vereinfachung dar. Die Anforderung der unbedingten Kontinuität im Sendebetrieb führte zu einer Vielzahl von Havarieszenarien, zu deren Bewältigung Infrastrukturelemente andere als ihre angestammten Rollen übernehmen und alternativ miteinander verschaltet werden können. Dies geht bis hin zum Rückfall auf Audio-/ Videokreuzschienen zur Weiterleitung des Materials innerhalb der Sendeanstalt. Die Expertise betrachtet ausschließlich den *Regelbetrieb* und beschränkt sich auch hier auf ausgewählte Anwendungsfälle.

4.1.6 Spezifische Anforderungen für die vertrauenswürdige und abgesicherte Langzeitarchivierung

Über die sich aus der Langzeitarchivierung ergebenden allgemeinen Anforderungen hinaus, stellt das Anwendungsszenario der öffentlich-rechtlichen Rundfunkanstalten weiterführende *Anforderungen* an die Beispielsysteme.

An das *Archiv* gestellte Anforderungen:

- Verwaltung von Roh- und Arbeitsmaterial für Arbeitsprozesse von Konsumenten und Produzenten
- Verwaltung von Metadaten für Arbeitsprozesse von Konsumenten und Produzenten
- Aufbewahrung von Archivalien verschiedenen Typs
- Führung von technischen und beschreibenden Metadaten, insbesondere auch Vorschauansicht und Schlüsselbilder
- Anpassbarkeit der Metadatenschemata an den Betreiber als auch an den Archivalientyp
- Strukturierte Untergliederung der Metadaten zeitlich als auch auf inhaltlicher Ebene
- Führung alternativer Fassung von Archivalien

- Bereitstellung und Verwaltung parallel genutzter, multipler Arbeitskopien für Konsumenten und Produzenten
- Zugriffskontrolle, Versionskontrolle und Rechtemanagement
- Schutz ausgewählter Archivalien vor Veränderung in einem von der Archivalie abhängigen Zeitraum
- Hohe Verfügbarkeit
- Uneingeschränkte Kontinuität des Geschäftsbetriebs des Betreibers
- Verwaltung temporär archivierter Fremdarchivalien
- Formatwechsel von Archivbeständen
- Vollständige Erschließung sämtlicher Inhalte der Produzenten

An den *Ingest* des Archivs gestellte Anforderungen:

- Häufige, zeitkritische Entgegennahme von Archivalien
- Häufige, zeitkritische Veränderung und Erweiterung von Archivalien
- Aufzeichnung von kontinuierlichen Medienströmen
- Hoher Datendurchsatz
- Übernahme von Archivalien aus anderen Archiven
- Redundante Zugriffswege für Havariebetrieb
- Anbindung an eine Vielzahl heterogener Produzentensysteme
- Einbindung in eine Vielzahl von Arbeitsprozessen
- Nutzung standardisierter Schnittstellen
- Entgegennahme zu archivierender, neuer digitaler Inhalte (Vorwärtsdigitalisierung)
- Digitalisierung vorhandener, physikalischer Archivalien (Rückwärtsdigitalisierung)

An den *Access* des Archivs gestellte Anforderungen:

- Häufige, zeitkritische Bereitstellung von Katalogdaten
- Häufige, zeitkritische respektive fristgerechte Bereitstellung von Archivalien
- Hoher Datendurchsatz
- Partielle Herstellung für die Bereitstellung lediglich kleiner Ausschnitte der Archivalien
- Abgabe von Archivalien an andere Archive
- Redundante Zugriffswege für Havariebetrieb
- Anbindung an eine Vielzahl heterogener Konsumentensysteme
- Einbindung in eine Vielzahl von Arbeitsprozessen
- Nutzung standardisierter Schnittstellen
- Bereitstellung von Vorschauansichten

An das *Archival Storage* des Archivs gestellte Anforderungen:

- Verantwortliche, sachgerechte Lagerung der Archivalien

An das *Preservation Management* des Archivs gestellte Anforderungen:

- Verantwortlicher langfristiger Erhalt der Archivalien
- Migration von Archivalien
- Erweiterung der Archivkapazitäten im laufenden Betrieb

An die *Administration* des Archivs gestellte Anforderungen:

- Selektive Löschung von Archivinhalten

- Redundante Zugriffswege für Havariebetrieb

An die *Produzenten* gestellte Anforderungen:

- Formale und inhaltliche Erschließung

An die *Konsumenten* gestellte Anforderungen:

Keine

4.2 Hochschul-Medienzentren

Die Konzeption und Strukturierung digitaler Langzeitarchive in Hochschul-Medienzentren ist in der Regel bemüht, das Rahmenwerk des OAIS-Referenzmodells [CCSDS02] abzubilden. Dies ist ebenso in der Verteilung der Akteure und ihren Rollen ersichtlich wie in der technischen Infrastruktur und den Informationsflüssen.

4.2.1 Akteure

In Bezug auf Langzeitarchivierungssysteme in Hochschul-Medienzentren lassen sich verschiedene konkrete Akteure benennen und in Anlehnung an das OAIS-Referenzmodell in die ein Archiv beeinflussenden Rollen Produzent, Konsument und Management aufteilen. Diese Akteure sind die Hochschuleinrichtung selbst, Bibliotheken, Rechenzentren, Archivdienstleister, Institute, Mitarbeiter, Studenten und externe Personen. Diese Auswahl erhebt keinen Anspruch auf Vollständigkeit. Akteure sind demnach zusammengefasste Nutzergruppen oder einzelne Personen. Akteure müssen nicht notwendigerweise der Hochschule angehören. Akteure können verschiedene Rollen gleichzeitig übernehmen. So ist die Hochschule gleichzeitig Produzent, Konsument und Management. In Abbildung 21 ist die Zuordnung dargestellt.

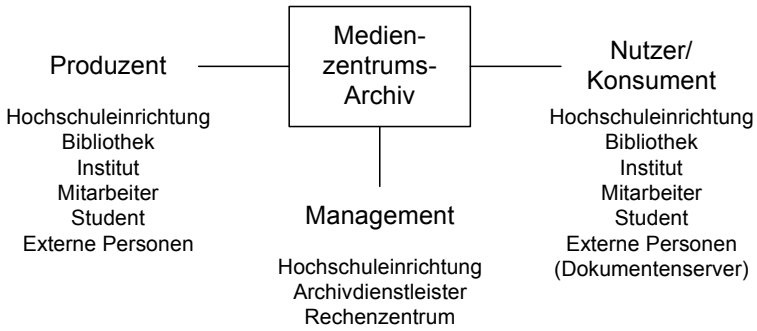

Abbildung 21: Akteure und Rollen in digitalen Langzeitarchiven in Hochschul-Medienzentren.

4.2.2 Architektur und Rollen

Die Organisation von Hochschul-Medienzentren teilt ein digitales Langzeitarchivsystem in die Hauptbereiche Erstellung/ Produktion, *Ingest, Access* und Archiv. Diesen Bereichen sind die zuvor genannten Akteure zugeordnet (Abbildung 22).

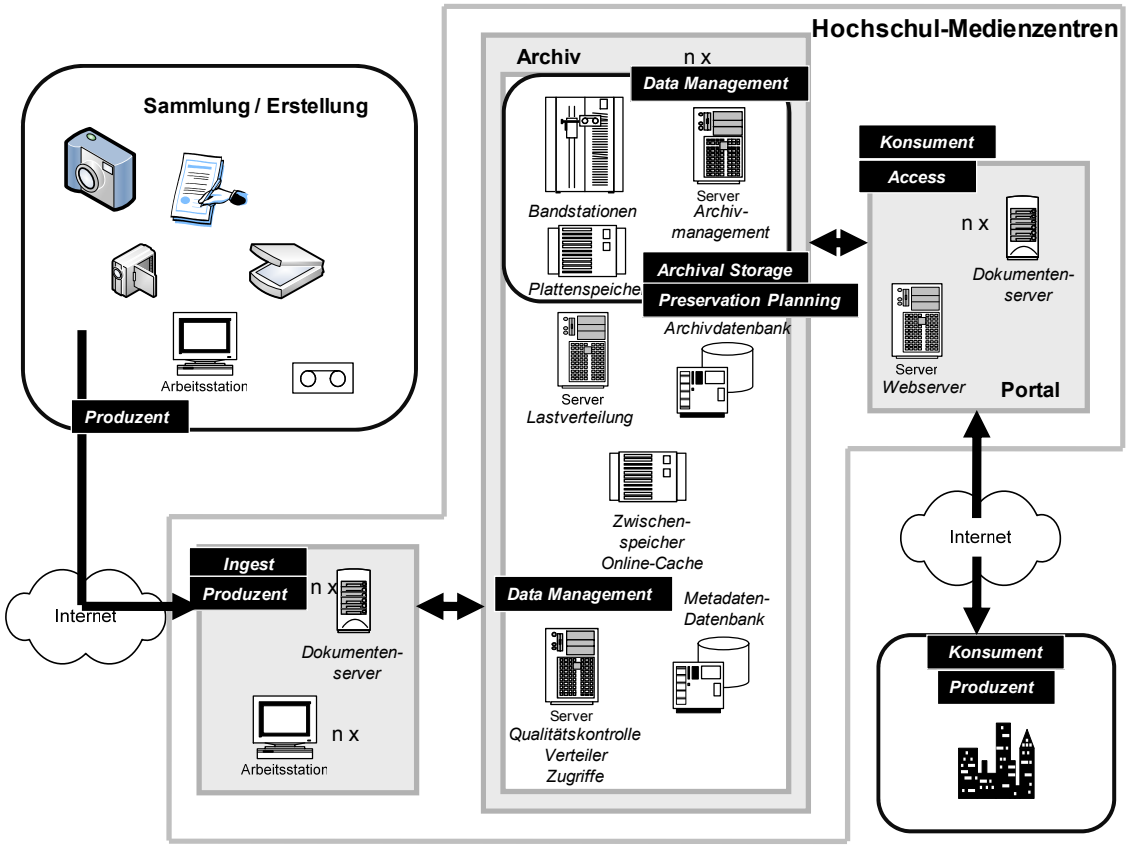

Abbildung 22: Auf die technische Infrastruktur bezogene Rollen und grobe Informationsflüsse in der Archivumgebung von Hochschul-Medienzentren.

Der Produzent erstellt, sammelt und produziert zu archivierende digitale Objekte. Diese werden dann in das Archiv eingeliefert, aufbereitet und erschlossen (*Ingest*) und werden dann über einen Zwischenspeicher in das Archiv eingespeist. Das Archiv muss dann seiner organisatorisch festgelegten Rolle des *Managements* gerecht werden und hat die Aufgabe des *Data Managements*, des *Archival Storage* und des *Preservation Planning*.

Die technische Infrastruktur der Hochschul-Medienzentren entspricht dem Referenzmodell. Das *Archival Storage* besteht aus Bandstationen und Plattenspeicher. Ein Zwischenspeicher ist meist vorweg geschaltet. Archivmanagement ist unterstützt durch Katalogdaten auf einem Datenbanksystem und regelt das Data Management. Aufgaben des Archivmanagements sind in dem Zusammenhang die Verwaltung der Bestandserhaltung und Bestandsbereitstellung. Hier sind Metadaten von entscheidender Bedeutung. Diese Metadaten sind im Ingest während der Bestandserschließung manuell oder automatisiert generiert.

4.2.3 Informationsflüsse

Die Informationsflüsse, die in dieser Studie Betrachtungsgegenstand sind, resultieren aus den Vorgängen:

- Bestandserweiterung,
- Bestandsbereitstellung und
- Bestandserhaltung.

Die Betrachtung der Informationsflüsse basiert auf einer abstrahierten technologischen Infrastruktur der Hochschul-Medienzentren.

Bestandserweiterung

Eingeliefert wird zuvor produziertes digitales Material, wie zuvor beschrieben u.a. digitalisierte Altbestände oder elektronisch generierte Information aller Medientypen. Der Informationsfluss zur

Bestandserweiterung ist in Abbildung 23 aufgezeigt. Die einzelnen Prozessschritte des Ablaufs sind nachfolgend in Tabelle 7 aufgeführt und kurz beschrieben.

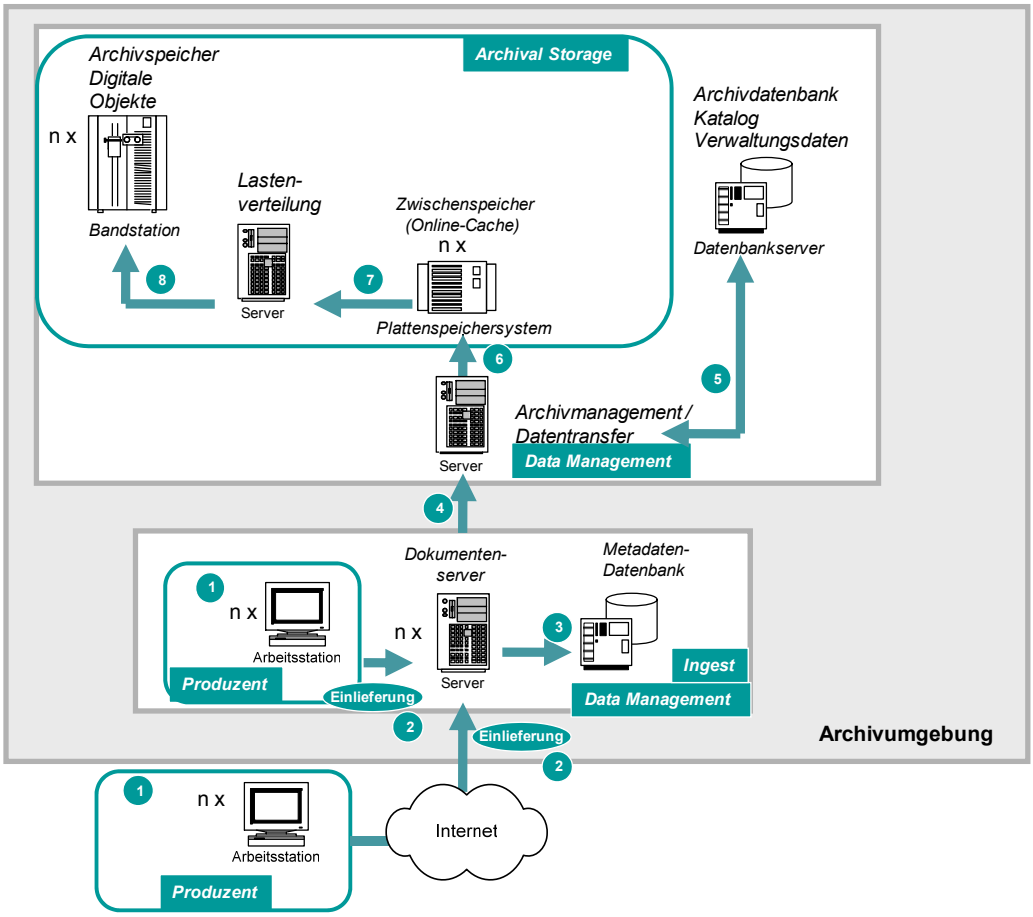

Abbildung 23: Rollen und Informationsflüsse bei einer Bestandserweiterung im abstrahierten Langzeitarchivierungssystem der Hochschul-Medienzentren.

Tabelle 7: Prozessablauf bei einer Bestandserweiterung im abstrahierten Langzeitarchivierungssystem des Szenarios Hochschul-Medienzentren.

Prozessschritt Nr.	Vorgang
1.	Erstellen des zu archivierenden Objektes.
2.	Einliefern des Objektes.
3.	Erschließung (manuell/ automatisiert) des Archivobjektes und Generierung von Metadaten, die dann in eine Datenbank eingepflegt und je nach Archiv auch mit dem Objekt gespeichert werden. Anschließend verbleibt das Archivobjekt zunächst auf dem Dokumentenserver.
4.	Übergabe des Archivobjektes an den Archivserver.
5.	Katalogisierung und Aufnahme von Identifier für Archivobjekt in eine Archivdatenbank, um Objekt im Archiv wieder zu finden.
6.	Unmittelbare Weitergabe des Archivobjektes an einen Zwischenspeicher. Dort wird das Objekt eine bestimmte Zeit liegen bevor es im nächsten Schritt in den eigentlichen Archivspeicher eingespeist wird. Sofern vorhanden wird ausgehend vom Zwischenspeicher-Inhalt eine Kopie an einem zweiten Archiv-Standort gesendet. Ggf. Erstellung einer kompletten Spiegelung neuer Archivalien durch das Archivmanagement.

7.	Weitergabe des Archivobjektes an einen Lastenverteilungsserver.
8.	Unmittelbare Weiterleitung des Archivobjektes vom Lastenverteilungsserver an den eigentlichen Archivspeicher.

Bestandsbereitstellung

Einer Recherche und Anfrage zur Folge geschieht die Auslieferung der angefragten Archivobjekte an autorisierte Konsumenten bzw. wird zur Ansicht bereitgestellt. Der Informationsfluss zur Bestandsbereitstellung ist in Abbildung 24 aufgezeigt. Die einzelnen Prozessschritte des Ablaufs sind nachfolgend in Tabelle 8 aufgeführt und kurz beschrieben.

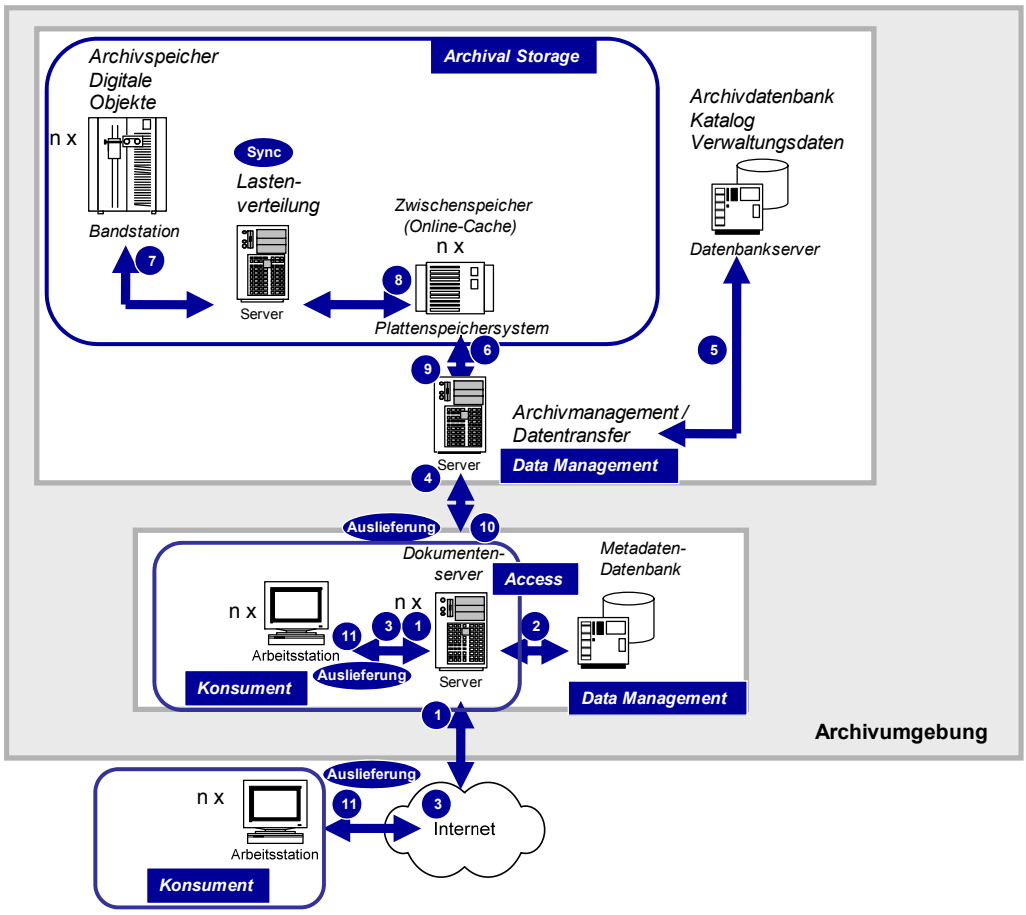

Abbildung 24: Rollen und Informationsflüsse bei einer Bestandsbreitstellung im abstrahierten Langzeitarchivierungssystem der Hochschul-Medienzentren.

Tabelle 8: Prozessablauf bei einer Bestandsbereitstellung im abstrahierten Langzeitarchivierungssystem des Szenarios Hochschul-Medienzentren.

Prozessschritt Nr.	Vorgang
1.	Recherche im Materialbestand über eine Arbeitsstation oder einen Online-Katalog. Stellen von Anfragen an das Archivsystem.
2.	Ermittlung zutreffender Katalogeinträge und Metadaten und Übermittlung der Ergebnismenge an die Arbeitsstation des anfragenden Konsumenten.
3.	Auswahl und Anforderung von Archivobjekt. Authentifizierung des Anfragenden Konsumenten seitens des Archivs. Liegt das Material auf dem Dokumentenserver wird es bei einer erfolgreichen Authentifizierung des Konsumenten an diesen aus-

	geliefert.
4.	Sofern das Material nicht bereits auf dem zentralen Speicher, dem Dokumentenserver vorliegt, Weiterleitung als Materialanfrage an den Server des Archiv-Managements.
5.	Kommunikation mit der Archivdatenbank, um das angeforderte Archivobjekt zu orten.
6.	Weiterleitung der Anfrage an den Zwischenspeicher, wenn Objekt dort geortet wurde.
7.	Wenn Archivobjekt im Archivspeicher geortet wurde, Weiterleitung der Anfrage dorthin.
8.	Bergung des Objektes vom Datenband und Überspielung in den Zwischenspeicher des Archivs. Ein Lastverteilmechanismus erlaubt die gleichzeitige Wiederherstellung aus gespiegelten Archivstandorten.
9.	Datentransfer des Archivobjektes über den Archivmanagement-Server.
10.	Auslieferung des Archivobjektes an den lokalen Dokumentenserver.
11.	Auslieferung zum Bereitstellungstermin an das ausgewählte Zielsystem.

Bestandserhaltung

Hochschul-Medienzentren stellen in ihren Systemen zur digitalen Langzeitarchivierung Mechanismen zur Bestandserhaltung zur Verfügung, auch wenn hier ein erheblicher Handlungsbedarf abzusehen ist. Der Informationsfluss zur Bestandserhaltung ist in Abbildung 25 aufgezeigt. Die einzelnen Prozessschritte des Ablaufs sind nachfolgend in Tabelle 9 aufgeführt und kurz beschrieben.

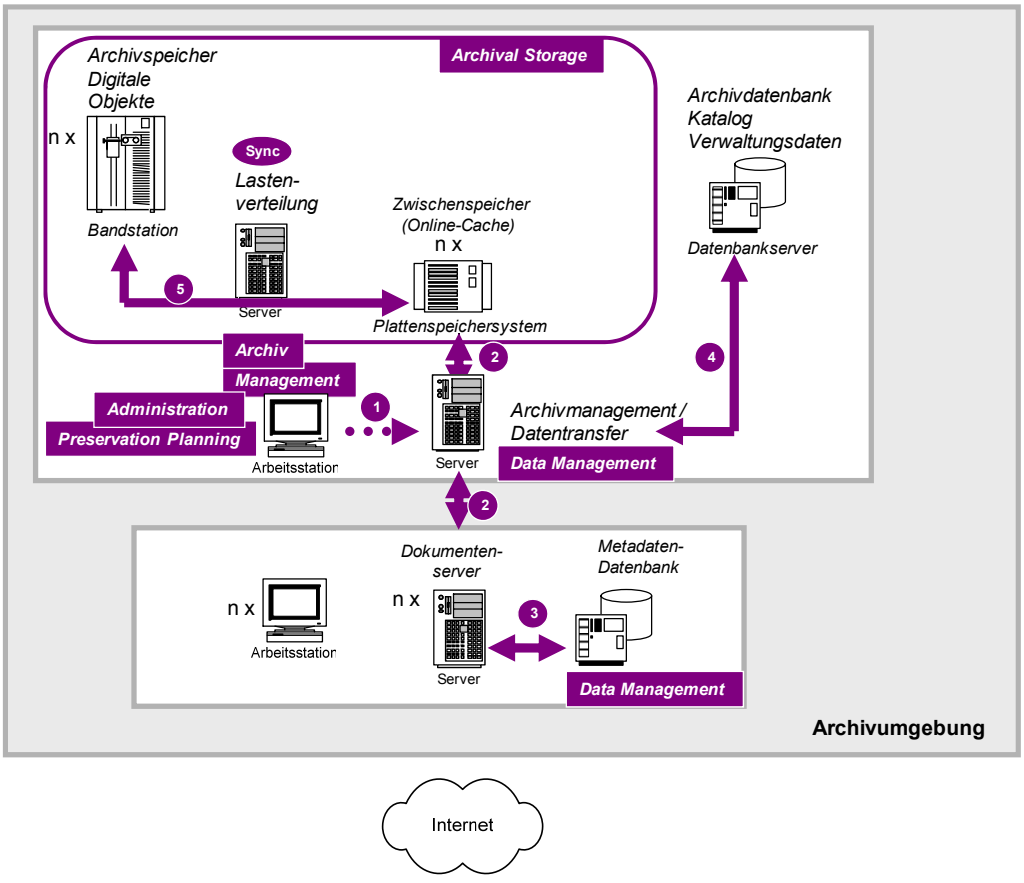

Abbildung 25: Rollen und Informationsflüsse bei einer Bestandserhaltung im abstrahierten Langzeitarchivierungssystem der Hochschul-Medienzentren.

Tabelle 9: Maßnahmen bei der Bestandserhaltung im abstrahierten System des Langzeitarchivierungssystem des Szenarios Hochschul-Medienzentren.

Prozessschritt/ Maßnahme Nr.	Vorgang
1.	Anweisung des Archivmanagements
2.	Weiterleitung der Anweisungen an Dokumentenserver und an eigentlichen Archivspeicher
3.	Überprüfung, Abgleich der Metadaten
4.	Datensicherung (*Backup*) der Katalogdaten.
5.	Datensicherung (*Backup*) Spiegelung der Archivinhalte und Katalogdaten, Banderneuerung, Transformierungen, Backup Umkopieren der Inhalte von als (potenziell) schadhaft erkannten Datenträgern.

4.2.4 Daten

Eine Zusammenführung des digitalen Objekts und der *Preservation Description Information (PDI)* zu einem *Information Package* findet im Bereich Hochschul-Medienzentren statt. Grundsätzlich lässt sich auch im Bereich der Hochschul-Medienzentren eine Unterscheidung nach *Submission Information Package (SIP)*, *Archival Information Package (AIP)* und *Dissemination Information Package (DIP)* nachvollziehen.

4.2.5 Annahmen über eine vertrauenswürdige und abgesicherte Langzeitarchivierung

Für die weitere Auswertungsarbeit werden in der Expertise die folgenden Annahmen über die Beispielsysteme der Hochschul-Medienzentren getroffen.

Es existiert eine zentrale Instanz (Archivmanagement) welche für die Verwaltung der Inhalte und Metadaten zuständig ist. Die Darstellungen der technischen Infrastruktur, Architektur und Informationsflüsse stellen eine Vereinfachung und Verallgemeinerung dar. Einzelne spezifische Schritte konnten nicht im Detail erhoben werden.

4.2.6 Spezifische Anforderungen für die vertrauenswürdige und abgesicherte Langzeitarchivierung

Über die sich aus der Langzeitarchivierung ergebenden allgemeinen Anforderungen hinaus stellt das Anwendungsszenario der Hochschul-Medienzentren weiterführende spezifische Anforderungen an die vertrauenswürdige und abgesicherte Langzeitarchivierung. Im Folgenden ist eine Auswahl aufgelistet, die keinen Anspruch auf Vollständigkeit erhebt. Die hier aufgestellten spezifischen Anforderungen ergeben sich aus der Systemabstraktion und den organisatorischen Verantwortlichkeiten.

An das *Archiv und Management* gestellte Anforderungen

- Organisatorische Hauptverantwortung
- Verwaltung, Erweiterung und Bereitstellung von Archivobjekten
- Verwaltung von Metadaten für die Wiederauffindbarkeit der Archivobjekte
- Aufbewahrung von Archivobjekten verschiedener Medientypen
- Verwaltung der Systemkomponenten und Ressourcen
- Anpassbarkeit der Komponenten
- Anpassbarkeit der Metadatenschemata
- Strukturierte Aufbereitung der Metadaten

- Bereitstellung von Sicherungsfassungen (*Backups*) von Archivobjekten
- Bereitstellung von Kopien
- Zugriffskontrolle
- Versionskontrolle
- Rechtemanagement (Berücksichtigung der aktuellen Gesetzeslage und Vorgaben)
- Schutz der Integrität und Authentizität
  - des Archivobjektes
  - der Systemkomponenten und Ressourcen (Hardware und Software) einschließlich
    - Speichermedium
    - Netzwerk
    - Darstellungsanwendung
- Hohe Verfügbarkeit
- Uneingeschränkte Kontinuität der Bereitstellung an autorisierte Konsumenten
- Bereitstellung von Schnittstellen zu Austauschzwecken mit anderen Archiven
- Formatwechsel von Archivbeständen
- Vollständige und einheitliche Erschließung der Archivobjekte
- Sicherung der Authentizität, Integrität, Verfügbarkeit, Vertraulichkeit und Nachweisbarkeit
- Entscheidung über Strategien zur Bestandserhaltung

An den *Ingest* des Archivs gestellte Anforderungen

- Anpassbarkeit der Oberfläche
- Nutzung standardisierter Schnittstellen
- Vollständige und einheitliche Erschließung der Archivobjekte
- Entgegennahme zu archivierender neuer digitaler Objekte
- Entgegennahme digitalisierter analog physikalischer Objekte
- Hoher Datendurchsatz
- Generierung von Metadaten
- Einbindung in eine Mehrzahl von Prozessen
- Überprüfung des eingelieferten Objekts auf Integrität und Authentizität

An den *Access* des Archivs gestellte Anforderungen

- Häufige und kontinuierliche Bereitstellung von Katalogdaten
- Häufige und zeitkritische Bereitstellung von Archivobjekten
- Hoher Datendurchsatz
- Authentifizierung der Konsumenten
- Redundante Zugriffswege bei Ausfall
- Sichere Anbindung in eine Vielzahl heterogener Konsumentensysteme
- Einbindung in eine Vielzahl von Arbeitsprozessen
- Sichere Auslieferung des integeren und authentischen Objektes bzw. der Objektkopie an authentifizierten Konsumenten
- Sicherung der Verfügbarkeit, Vertraulichkeit und Nachweisbarkeit

An das *Archival Storage* des Archivs gestellte Anforderungen

- Verantwortliche, sachgerechte und sichere Lagerung der Archivobjekte
- Verantwortliche, sachgerechte und sichere Lagerung der Datenbanken

- Synchronisation der Speichermedien

An das *Preservation Management* gestellte Anforderungen

- Verantwortlicher, langfristiger und abgesicherter Erhalt der Archivobjekte
- Bereitstellung angemessener Bestandserhaltungsmaßnahmen wie Migration oder Emulation
- Erweiterung der Archivkapazitäten im laufenden Betrieb
- Anpassbarkeit
- Lebenszyklusmanagement der Archivobjekte

An die *Administration* des Archivs gestellte Anforderungen

- Einbindung der aktuellen Gesetzeslage und Vorgaben in Rechtemanagement
- Selektive Löschung
- Redundante Zugriffswege bei Ausfall
- Steuerung des Archivs
- Umsetzung, Weiterleitung und Verteilung der Verantwortungen

An die *Produzenten* gestellte Anforderungen

- Sammlung und Akquirierung der potentiellen Archivobjekte
- Formale und inhaltliche Erschließung

An die *Konsumenten* gestellte Anforderungen

- Formell: Verantwortungsvoller Umgang mit ausgelieferten Archivobjekten

5 Herangehensweisen zur Integration von Sicherheitstechnologien für eine vertrauenswürdige und abgesicherte Langzeitarchivierung multimedialer Inhalte

In diesem Kapitel werden Herangehensweisen zur Integration von Sicherheitstechnologien für eine vertrauenswürdige und abgesicherte Langzeitarchivierung multimedialer Inhalte beschrieben. Für die Erstellung einer Methodologie werden dazu zunächst die *Sicherheitsaspekte* allgemein dargestellt und die Auswirkungen von Verletzungen der Sicherheitsaspekte werden beschrieben. *Bedrohungen* werden aufgezeigt und Angriffe dargelegt. Zudem werden Sicherheitsrichtlinien und IT-Sicherheitsmanagement kurz erläutert. Folglich wird die Aufgabe von Sicherheitsmechanismen beschrieben und beispielhafte Sicherheitsmechanismen werden benannt. Im Zusammenhang der Integration von Sicherheitstechnologien für eine vertrauenswürdige und abgesicherte Langzeitarchivierung digitaler Information wird zusätzlich auf die *Common Criteria* näher eingegangen, Vertrauen und Vertrauenswürdigkeit werden abgegrenzt und Dokumentation und Transparenz werden gegenübergestellt.

Aufbauend auf der allgemeinen Darstellung werden die Sicherheitsaspekte im *Kontext der Langzeitarchivierung* beschrieben und die Einsatzmöglichkeiten von Sicherheitstechnologien und deren Eignung zur vertrauenswürdigen und abgesicherten Langzeitarchivierung wird evaluiert. Dazu wird zunächst die *Bedeutung* der fünf Sicherheitsaspekte für digitale Langzeitarchive dargelegt.

Ausgehend von den Sicherheitsaspekten werden darauf aufbauend die Sicherheitstechnologien den zuvor erhobenen *Anforderungen* (Soll) an vertrauenswürdige und abgesicherte Langzeitarchive zugeordnet.

Für die Reflektion der Sicherheitstechnologien in Bezug auf die organisatorischen und technischen Rahmenbedingungen innerhalb der betrachteten Szenarien werden die Anforderungen mit den zugeordneten möglichen einsetzbaren Sicherheitstechnologien mit der Erhebung des *Ist-Zustandes* und des Handlungsbedarfs für Hochschul-Medienzentren und für Rundfunkanstalten vervollständigt. Die Erhebung des Ist-Zustandes beruht auf existierenden und uns zur Verfügung gestellten Studien und weiterführenden Interviews. Auf den Handlungsbedarf wird in Kapitel 6 näher eingegangen. Abschließend erfolgt eine Validierung der Einsetzbarkeit in praktischen Beispielen.

5.1 Allgemeine Einsatzmöglichkeiten von Sicherheitstechnologien und deren Eignung zur vertrauenswürdigen und abgesicherten Langzeitarchivierung multimedialer Inhalte

In diesem Abschnitt werden allgemeine Einsatzmöglichkeiten von Sicherheitstechnologien und deren Eignung zur vertrauenswürdigen und abgesicherten Langzeitarchivierung multimedialer Inhalte erhoben und diskutiert.

Einführend werden die Sicherheitsaspekte im Allgemeinen aufgeführt und die Auswirkungen von Verletzungen der Sicherheitsaspekte werden dargelegt. Dies beinhaltet eine Beschreibung der Bedrohungen und Angriffe, der Sicherheitsrichtlinien und des IT-Sicherheitsmanagements. Die generellen Aufgaben von Sicherheitsmechanismen werden definiert und beispielhafte Sicherheits-

mechanismen werden benannt. *Common Criteria* wird aufgeführt, Vertrauen und Vertrauenswürdigkeit werden abgegrenzt und Dokumentation und Transparenz werden gegenübergestellt.

Im Folgenden wird die Bedeutung der fünf Sicherheitsaspekte für digitale Langzeitarchive dargelegt. Sicherheitstechnologien werden den zuvor erhobenen Anforderungen (Soll) an vertrauenswürdige und abgesicherte Langzeitarchive zugeordnet und ihre Eignung wird ausgewertet.

5.1.1 Digitale Langzeitarchivierung und IT-Sicherheit

Das *Ziel von IT-Sicherheit* in jedem IT-System [Eck06] ist die Begrenzung bzw. Minimierung der Verwundbarkeit von Werten und Ressourcen, d.h. von Schwachstellen und dem dadurch resultierenden Gefahrenpotential. (ISO IS 7498/2 Security Architecture). Digitale Information ist in Bitströmen (digitale Zeichenströme) repräsentiert. Ein *Bitstrom* hat den Vorteil einheitlich gespeichert werden zu können und kann damit nahezu perfekt kopiert werden. Gleichzeitig ist die digitale Information gegenüber der analogen Information dadurch jedoch angreifbarer in ihrer Integrität und Authentizität, da Manipulationen bzw. Veränderungen schneller, einfacher und unbemerkter geschehen können. Des Weiteren können durch den *Wechsel ihrer Repräsentationsform* (Formats), was in digitalen Systemen und gerade in Systemen zur Langzeitarchivierung bedingt durch Konversionen und Migrationen zur Erhaltung der Information wiederholt geschieht, so genannte Medienbrüche verursacht werden. Information kann verloren gehen und ist somit nicht mehr integer oder authentisch. Zudem wird ein so genanntes Abspielsystem benötigt, um digitale Information anzuzeigen und dem Menschen zugänglich zu machen. Auch hier können Medienbrüche entstehen, nämlich dann, wenn ein solches System zur Anzeige nicht verfügbar oder veraltet ist. Basierend auf diesen Problemen werden im Kontext von IT-Sicherheit immer die folgenden *Sicherheitsaspekte* betrachtet.

5.1.2 Sicherheitsaspekte allgemein

Sicherheitsaspekte sind Charakteristiken, die ein sicheres digitales IT-System aufweisen sollte. Hierzu zählen Verfügbarkeit, Vertraulichkeit, Integrität, Authentizität und Nachweisbarkeit bzw. Nicht-Abstreitbarkeit [Dit04].

Verfügbarkeit (*Availability*)

Verfügbarkeit bezeichnet die Sicherheitseigenschaft, dass Ressourcen, Information, Dienste und Geräte auf Abruf für autorisierte Parteien/ Benutzer bereitstehen und von diesen zu jedem vereinbarten bzw. geplanten Zeitpunkt ohne Einschränkungen genutzt werden können. Ein verfügbares System muss dabei vor unautorisierten Zugriffen geschützt sein.

„*Sei X eine Menge von Personen und sei I eine Ressource. Dann hat I die Eigenschaft der Verfügbarkeit in Bezug auf X, wenn alle Mitglieder von X auf I zugreifen können*“ [Bis03].

Vertraulichkeit (*Confidentiality*)

Vertraulichkeit bezeichnet den Schutz von Information vor Zugriff durch unbefugte Personen und besagt, dass nur autorisierte Personen Zugriff bzw. Zugang zu bestimmten Information haben bzw. dass Personen nicht unautorisiert Kenntnis von Information erlangen. Folglich beinhaltet Vertraulichkeit den Begriff der Privatsphäre.

„Sei X eine Menge von Personen und sei I eine gewisse Information. Dann hat I die Eigenschaft der Vertraulichkeit in Bezug auf X, wenn kein Mitglied von X Information über I erhalten kann“ [Bis03].

Integrität (*Integrity*)

Integrität bezeichnet die Unversehrtheit von Information oder Ressourcen und damit den Schutz vor zielgerichteten Manipulationen oder unbeabsichtigten Veränderungen. Integrität bedeutet die Vollständigkeit, Konsistenz und Genauigkeit/ Korrektheit von Information. So schließt die Integrität den Schutz von Information vor unberechtigten Veränderungen ein.

„Sei X eine Menge von Personen und sei I entweder eine Information oder eine Ressource. Dann hat I die Eigenschaft der Integrität in Bezug auf X, wenn alle Mitglieder von X dieser Information oder Ressource vertrauen“ [Bis03].

Authentizität (*Authenticity*)

Authentizität bezeichnet die Echtheit von Information oder von Personen (Entitäten). Authentizität beinhaltet die sichere Bestimmung (den Nachweis) des echten Ursprungs von Information und deren Urhebern sowie der autorisierten Kommunikationspartner. Authentizität von Information ist folglich die Bestimmung der Originalität. Authentizität von Entitäten bezieht sich auf die Sicherstellung der Echtheit von Sender und Empfänger (Kommunikationspartner) und bezeichnet den sicheren Nachweis, dass eine Person (Entität) richtig identifiziert/ verifiziert ist bzw. dass eine Nachricht so angekommen ist, wie sie gesendet worden ist.

Differenzierung Authentizität und Integrität

Integrität verweist auf die Veränderung einer Information während Authentizität den Identitätsnachweis, den Nachweis des Ursprungs (Origin) einer Information, also das Original bezeichnet.

Die Aspekte Integrität und Authentizität werden auch im nestor-Kriterienkatalog [Nes06] getrennt behandelt. Dies ist im Teil B „Umgang mit Objekten“ aufgeführt. Dieser Teil beinhaltet die vom digitalen Langzeitarchiv analysierten Ziele und Strategien und spezifizierten objektbezogenen Anforderungen für den Umgang mit digitalen Objekten während des Lebenszyklus der Objekte im digitalen Langzeitarchiv. Dies entspricht den Hauptphasen des OAIS-Referenzmodells [CCSDS02] mit den Prozessen (in OAIS als funktionalen Entitäten bezeichnet) Aufnahme (*Ingest*), Archivablage (*Archival Storage*, inklusive Umsetzung der Langzeiterhaltungsmaßnahmen *Preservation Planning*) und Nutzung (*Access*).

Im Punkt 6 im Teil B des Kriterienkatalogs ist Integrität wie folgt beschrieben:

> *„Das digitale Langzeitarchiv stellt die Integrität der digitalen Objekte auf allen Stufen der Verarbeitung sicher. Unter Integrität wird hier erstens die Vollständigkeit der digitalen Objekte sowie zweitens der Ausschluss unbeabsichtigter Modifikationen im Sinne der Erhaltungsregeln [(herbeigeführt durch menschliche Akteure (böswillig oder irrtümlich) oder durch technische Unvollkommenheiten sowie Beschädigung oder Entwendung von technischer Infrastruktur)] verstanden. Maßstab für die Integrität sind die als erhaltenswert definierten Eigenschaften eines digitalen Objekts (Punkt 9.2).“*

Im Punkt 7 im Teil B des Kriterienkatalogs sind zur Authentizität folgende Ausführungen zu finden:

> *„Das digitale Langzeitarchiv stellt die Authentizität der digitalen Objekte auf allen Stufen der Verarbeitung sicher. Authentizität bedeutet hier, dass das Objekt das darstellt, was es vorgibt darzustellen. Das dLZA dokumentiert, wenn bei einem Objekt die Authentizität nicht festgestellt werden kann. Das dLZA betreibt ein für den Erhalt der Authentizität geeignetes Datenmanagement für die Prozesse Aufnahme (Ingest), Archivablage (Archival Storage) und Nutzung (Access). Dieses stellt insbesondere die Dokumentation aller Veränderungen an den Objekten (inkl. Metadaten) sicher (Punkt 12.4).“*

In Bezug auf digitale Langzeitarchivierungssysteme kann dies an unterschiedlichen Beispielen verdeutlicht werden. So muss sich z.B. eine Migration einer Information im Normalfall nicht zwangsläufig auf die Integrität der Information auswirken, d.h. nicht notwendigerweise zu einem Informationsverlust führen. Es wird zwar möglicherweise das Daten-Modell, das Daten-Schema oder das Daten-Format der Information geändert, also die syntaktische Darstellungsweise, aber der semantische Inhalt bleibt annähernd gleich. Je nach Repräsentationform in welche die Information migriert wird, kann dies aber ebenso auch Auswirkungen auf die Integrität haben, nämlich dann, wenn eine z.B. so genannte verlustbehaftete Kompression (*Lossy Compression*) während der Transformation stattfindet. Dann können die Informationsänderungen zum Integritätsverlust führen. Die Authentizität ist mit einer Migration in jedem Fall verloren, da die Originalität nicht mehr gegeben ist. In einem

digitalen Langzeitarchiv kann demnach nicht immer ein Schutz der Integrität und/ oder der Authentizität gewährleistet werden. Um dennoch vertrauenswürdig und abgesichert zu sein, müssen daher alle Veränderungen dokumentiert werden.

In diesem Zusammenhang sei auch auf das *DOMEA (Dokumentenmanagement und elektronische Archivierung im IT-gestützten Geschäftsgang) Konzept* [DOM05] für Dokumentenmanagement und elektronische Archivierung in der öffentlichen Verwaltung in Deutschland verwiesen, dessen wesentliches Ziel die Einführung der elektronischen Akte ist. Das DOMEA-Konzept liefert die Richtlinien für eine vollständige Überführung behördlicher Geschäftsprozesse, Vorgangsbearbeitungen und Archivierung in konforme IT-Prozesse. Da für das elektronische wie auch das Papierschriftgut die in Gesetzen, Geschäftsordnungen sowie Richtlinien und Vorschriften festgeschriebenen Anforderungen gleichermaßen gelten, müssen behördliche Unterlagen auch in elektronischer Form den Kriterien Vollständigkeit, Integrität und Authentizität, Zusammenfassung aufgabenbezogener und zusammengehöriger Schriftstücke, Nachvollziehbarkeit und Rechtmäßigkeit des Verwaltungshandelns genügen. Gleichermaßen müssen auch elektronische Akten eine transparente und nachvollziehbare Struktur aufweisen und sich in einen Kontext einordnen lassen. [DOM05]

Das DOMEA-Konzept ist eng verbunden mit der ISO Norm 15489, welche ein internationaler Standard ist, der in Deutschland unverändert als DIN-Norm übernommen wurde und Leitlinien zur Verwaltung von Schriftgut von öffentlichen und privaten Organisationen bietet. Mit der Norm soll ein Rahmen für die Verwaltung und Aufbewahrung von Unterlagen unabhängig von ihrer physischen Beschaffenheit und der logischen Struktur geschaffen werden. Gleichzeitig werden als Qualitätsstandard Regeln für transparente, nachvollziehbare Verwaltungs- oder Geschäftsvorgänge definiert. In der ISO Norm 15489 wird das so genannte *Records Management* (RM) definiert als Führungsaufgabe wahrzunehmende, effiziente und systematische Kontrolle und Durchführung der Erstellung, Entgegennahme, Aufbewahrung, Nutzung und Aussonderung von Schriftgut einschließlich der Vorgänge zur Erfassung und Aufbewahrung von Nachweisen und Information über Geschäftsabläufe und Transaktionen in Form von Akten.[24] Das *Records Management* bezieht neben dem eigentlichen Inhalt von Records (*Content*) auch den Entstehungszusammenhang (Kontext) ein. *Records* (Unterlagen bzw. Akten) bezeichnen alle, unabhängig vom Informationsträger, erstellten oder empfangenen geschäftsrelevanten Information zzgl. aller Hilfsmittel und Metadaten, die für das Verständnis einer Information und deren Nutzung notwendig sind.

Die Integrität eines *Records* ist im *Records Management* festgelegt als Abhängige von den drei Größen: Inhalt, Kontext und Struktur des originalen *Records*. Nichts anderes gilt für die Handhabung der digitalen Objekte innerhalb der digitalen Langzeitarchivierung. Die Authentizität eines *Records* ist im *Records Management* als entscheidende Größe definiert, um zu bestimmen ob eine Information das echte Record eines Ereignisses ist. So gibt ein *Record* beispielsweise den Nachweis, dass ein Vertrag zustande gekommen ist. In Umgebungen mit einem sehr hohen Datenaufkommen, wie in digitalen Systemen können sich verloren gegangene *Records* daher auf ein Einbüßen der Produktivität und erhöhte Kosten auswirken. So kann der Nachweis eines Ereignisses nicht mehr erbracht werden und die Authentizität ist nicht mehr gewährleistet.

Zusammenfassend lässt sich sagen, dass Integrität und Authentizität durchgehend als Kernpunkte im Kontext der digitalen Langzeitarchivierung betrachtet werden müssen.

Nachweisbarkeit/ Verbindlichkeit/ Nicht-Abstreitbarkeit (*Non-Repudiation*)

Die Nachweisbarkeit gibt an, dass sichergestellt werden kann, ob ein bestimmtes Ereignis im digitalen IT-System verbindlich stattfand und nicht abgestritten werden kann. Sie dient somit der Aufklärung von Ursachen für Informationsverletzungen. In Bezug auf den Austausch von Nachrichten bedeutet die Nachweisbarkeit (Verbindlichkeit/ Nicht-Abstreitbarkeit), dass der Absender einer Nachricht in der Lage ist, den Empfang der Nachricht durch den Empfänger nachzuweisen. Weiterhin kann der Empfänger einer Nachricht nachweisen, dass der Absender tatsächlich die Nachricht gesendet hat [BaAr05]. Das Versenden bzw. Empfangen von Nachrichten ist mittels authentisch festgestellter Personen unabstreitbar geschützt.

[24] http://www.iso.org/iso/catalogue_detail?csnumber=31908

Nachweisbarkeit (Verbindlichkeit/ Nicht-Abstreitbarkeit) bezieht sich auf die Einhaltung der vorher genannten Sicherheitsaspekte Verfügbarkeit, Vertraulichkeit, Integrität und Authentizität. Der Ursprung von Information und deren Urheber kann unabstreitbar, verbindlich und fälschungssicher nachgewiesen werden, ebenso die Identität einer Person. Dies kann von authentisch festgestellten Personen verifiziert werden.

5.1.3 Digitales Langzeitarchiv – Auswirkungen von Verletzungen der IT-Sicherheitsaspekte

Die Auswirkungen von Verletzungen der IT-Sicherheitsaspekte vollziehen sich auf zwei Ebenen: funktional oder strukturell. Dies gilt sowohl für den Archivinhalt als auch die Archivstruktur.

Funktionale Auswirkungen beziehen sich lediglich auf ein betroffenes Element, dessen IT-Sicherheitsaspekte verletzt sind, nicht aber auf andere Elemente im Archivsystem. Das Element ist in seiner eigenen Funktionalität beeinträchtigt. Betrifft dies z.B. die Integrität eines digitalen Objektes, so ist dieses Objekt nicht mehr unversehrt bzw. vollständig, es ist verändert, was aber keinerlei Auswirkungen auf andere Objekte oder Elemente im Archivsystem hat. Der Schaden begrenzt sich in diesem Fall auf das betroffene Element.

Strukturelle Auswirkungen dagegen beziehen sich auf das gesamte Netzwerk eines Archivsystems einschließlich aller darin enthaltenen Komponenten. Als Strategie für vertrauenswürdige Langzeitarchivierungssysteme werden Archivsysteme zunehmend durch z.B. LAN, WAN, Internet oder moderne Grid-Technologien vernetzt. Zu archivierende Objekte unterliegen bei ihrer Erstellung bzw. Aufbereitung einem so genannten Erschließungsprozess, d.h. zusätzliche Metadaten werden generiert. Metadaten werden zum Teil unabhängig vom eigentlichen Objekt abgelegt. Die Objekte und deren Metadaten sind auf unterschiedlichen, miteinander verbundenen Rechnern abgelegt, werden auf Anfrage abgerufen und zu einem Multimedia-Objekt komponiert.

Strukturelle Auswirkungen von Verletzungen von IT-Sicherheitsaspekten resultieren entweder aus einer vorhergehenden funktionalen Schadensauswirkung, also der Beeinträchtigung der Funktionalität eines Elements des Netzwerkes, können aber auch auftreten, ohne dass das auslösende Element selbst betroffen ist, sondern nur als Träger fungiert. Bei einer strukturellen Auswirkung sind mehrer Komponenten im Netzwerk in ihrer Funktion beeinträchtigt. Im schlimmsten Fall (*Worst Case*) ist das gesamte Netzwerk funktionsunfähig. Die folgenden Abbildungen demonstrieren beispielhaft und verallgemeinert die unterschiedlichen Auswirkungen (funktional und strukturell) innerhalb von Netzwerken.

Funktionale Auswirkungen

Funktionale Auswirkungen werden in Abbildung 26 dargestellt. Die Auswirkungen beschränken sich auf ein einzelnes Element (schwarz dargestellt) des Netzwerks (links) oder mehrere einzelne, voneinander unabhängige Elemente des Netzwerks (rechts). Andere Elemente des Netzwerks sind nicht betroffen. Das betroffene Element muss nicht notwendigerweise Komponente eines Netzwerks sein. In einem Archivsystem wären in diesem Fall lediglich ein oder mehrere einzelne Archivobjekte betroffen, nicht aber die Funktionalität des Archivsystems.

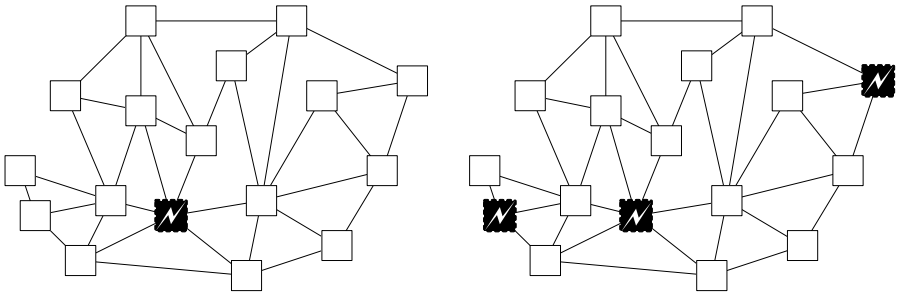

Abbildung 26: Funktionale Auswirkungen.

Strukturelle Auswirkungen

Strukturelle Auswirkungen werden in Abbildung 27 dargestellt. Ausgehend von einem primär betroffenen Element oder Trägerelement (schwarz) breiten sich die Auswirkungen auf andere Komponenten des Netzwerks (schwarz umrandet) aus. Dabei müssen nicht notwendigerweise alle anderen Netzwerkkomponenten betroffen sein (links). Dies kann z.B. zur Folge haben, dass bestimmte Komponenten vom Netzwerk isoliert werden (Knoten „A"). Das isolierte Element kann folglich mit keiner anderen Komponente des Netzwerks kommunizieren, da es nur Verbindung zu beschädigten Komponenten aufweist und diese nicht funktionsfähig sind. Ein Archivsystem würde in diesem Fall je nach Schadensauswirkung zum Teil funktionieren, wäre zum Großteil aber nur sehr eingeschränkt funktionsfähig. Auf der rechten Seite ist der schlimmste mögliche Fall demonstriert, nämlich dass das komplette Netzwerk stillgelegt ist. Ein Archivsystem wäre in diesem Fall komplett nicht verfügbar.

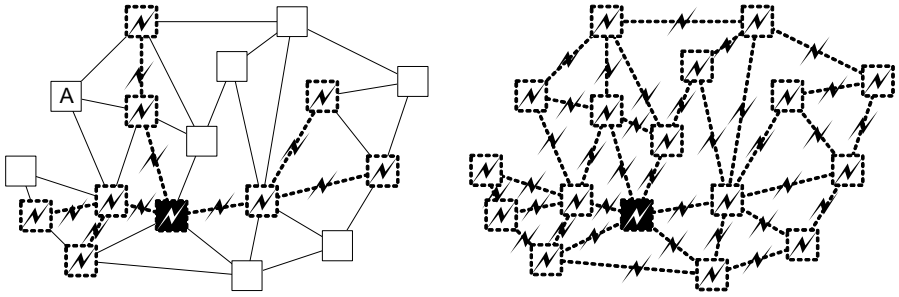

Abbildung 27: Strukturelle Auswirkungen.

Die strukturellen Auswirkungen können auch zeitlich versetzt auftreten, nämlich dann wenn ein Element unerkannt geschädigt bzw. kompromittiert ist. So wird beispielsweise ein mit einem Schadprogramm (Virus, Trojanisches Pferd, Wurm, usw.) befallenes digitales Objekt in ein digitales Langzeitarchiv aufgenommen, ohne dass der Virus detektiert worden ist, weil z.B. kein Antivirenprogramm aktiv ist oder dieses das Schadprogramm nicht erkannt hat. Das Schadprogramm ist nun so programmiert, dass es erst nach einer bestimmten Zeit aktiv wird. So wird das vermeintlich saubere Objekt zum Trägerelement, von dem der strukturelle Schaden ausgeht und sich auf andere Elemente im Archiv überträgt. Dies stellt eine der größten Bedrohungen im digitalen Langzeitarchiv dar.

Beispiel Virus

Als Beispiel sei hier die Archivierung eines digitalen Objekts aufgeführt, das mit einem Virus befallen ist. Dies stellt eine Verkettung von Auswirkungen dar. Zunächst ist davon auszugehen, dass das primär schadhafte Element, also das Objekt, in seiner Funktionalität eingeschränkt ist. Davon ausgehend verbreitet sich der Virus in dem ganzen Netzwerk und kann im schlimmsten Falle das gesamte Netzwerk befallen. Somit wären die IT-Sicherheitsaspekte verletzt und das Archivsystem nicht mehr vertrauenswürdig einsatzfähig.

5.1.4 Allgemeine Bedrohungen

Das Bundesamt für Sicherheit in der Informationstechnik (BSI) hat die **Gefährdungen** in fünf unterschiedliche Gefährdungskataloge [BSI06] aufgeteilt**:**

G 1: Höhere Gewalt

G 2: Organisatorische Mängel

G 3: Menschliche Fehlhandlungen

G 4: Technisches Versagen

G 5: Vorsätzliche Handlungen

Vorsätzliche Handlungen (Gefährdung G 5): Risiken und Angriffe

In digitalen Systemen, die durch das Internet oder Netzwerke verbunden sind, herrscht ein ständiger Informationsfluss und Austausch. Vernetzte digitale Kommunikationssysteme bieten eine breite

Angriffsfläche und bergen je nach Ziel und Intention des Angreifers unterschiedliche Risiken bzgl. der IT-Sicherheitsaspekte.

In diesem Zusammenhang soll zunächst zwischen *Safety* und *Security* unterschieden werden. *Safety* ist der Zustand sicher bzw. geschützt zu sein vor Gefahr, Schaden, Verletzungen verursacht durch zufällige, unbeabsichtigte und hauptsächlich unvorhersehbare Einflüsse und Einwirkungen. *Safety* beinhaltet dabei Mechanismen zur Schadensbehebung- und Vermeidung und Wiederherstellung zufälliger, unbeabsichtigter Schäden und Verletzungen. Zu solchen Gefahren zählen zum Beispiel Naturkatastrophen oder technische Störungen. In Eckert wird Safety definiert: „Unter Funktionssicherheit (*Safety*) eines Systems verstehen wir die Eigenschaft, dass die realisierte Ist-Funktionalität der Komponenten mit der spezifizierten Soll-Funktionalität übereinstimmt. Ein funktionssicheres System nimmt keine funktional unzulässigen Zustände an. Anders herum verstehen wir unter der Funktionssicherheit eines Systems, dass es unter allen (normalen) Betriebsbedingungen funktioniert.“ [Eck06]

Im Kontext der digitalen Langzeitarchivierung wird *Safety* in dieser Expertise weniger berücksichtigt. Vielmehr geht es darum, die *Security* zu betrachten, d.h. vorhersehbare und gezielte Angriffe vorzubeugen, sie zu identifizieren und den Schaden einzugrenzen bzw. die Information wiederherzustellen. Sicherheit (*Security*) kann als Risikomanagement-Prozess angesehen werden.

Angreifer entwickeln ständig neue Angriffsvorgehen (Aktivitäten) und Strategien, um sich unautorisierten Zugang zu Accounts, Information, Prozessen, Computersystemen, Netzwerken und Internetnetzwerken zu verschaffen, diese zu manipulieren und dadurch letzten Endes den dahinter stehenden Identitäten (Personen, Firmen, Regierung, usw.) Schaden zuzufügen. Es gibt jedoch gleichermaßen Angreifer mit einer schadensabwendenden Intention, nämlich solche, die im Auftrag einer Schadensbeurteilung vielmehr Schaden abwenden als welchen anrichten wollen.

Generell lässt sich sagen, dass ein Angreifer immer ein bestimmtes Endziel bzw. eine Absicht hat, einen Grund weswegen er den Angriff startet. Er greift auf bestimmte Werkzeuge zurück und nutzt Schwachstellen im System aus. Je nach Angriffsziel und dem damit verbundenen Ergebnis, das er erreichen will, geht der Angreifer unterschiedlich vor. Dies ist in Abbildung 28 schematisch und allgemein dargestellt.

Ein Ereignis ist eine Aktivität, eine Handlung, die an einem bestimmten Ziel/ Gegenstand ausgerichtet ist, mit der Absicht bzw. dem gewünschten Ergebnis, den Zustand des Ziels/ Gegenstands zu ändern. (IEEE96:373) Ein Benutzer führt die Aktivität Authentifizieren aus mit der Absicht, sich in seinen Account einzuloggen. Die Aktivität startet einen Prozess (Login-Programm), damit der Benutzer sich einloggen kann.

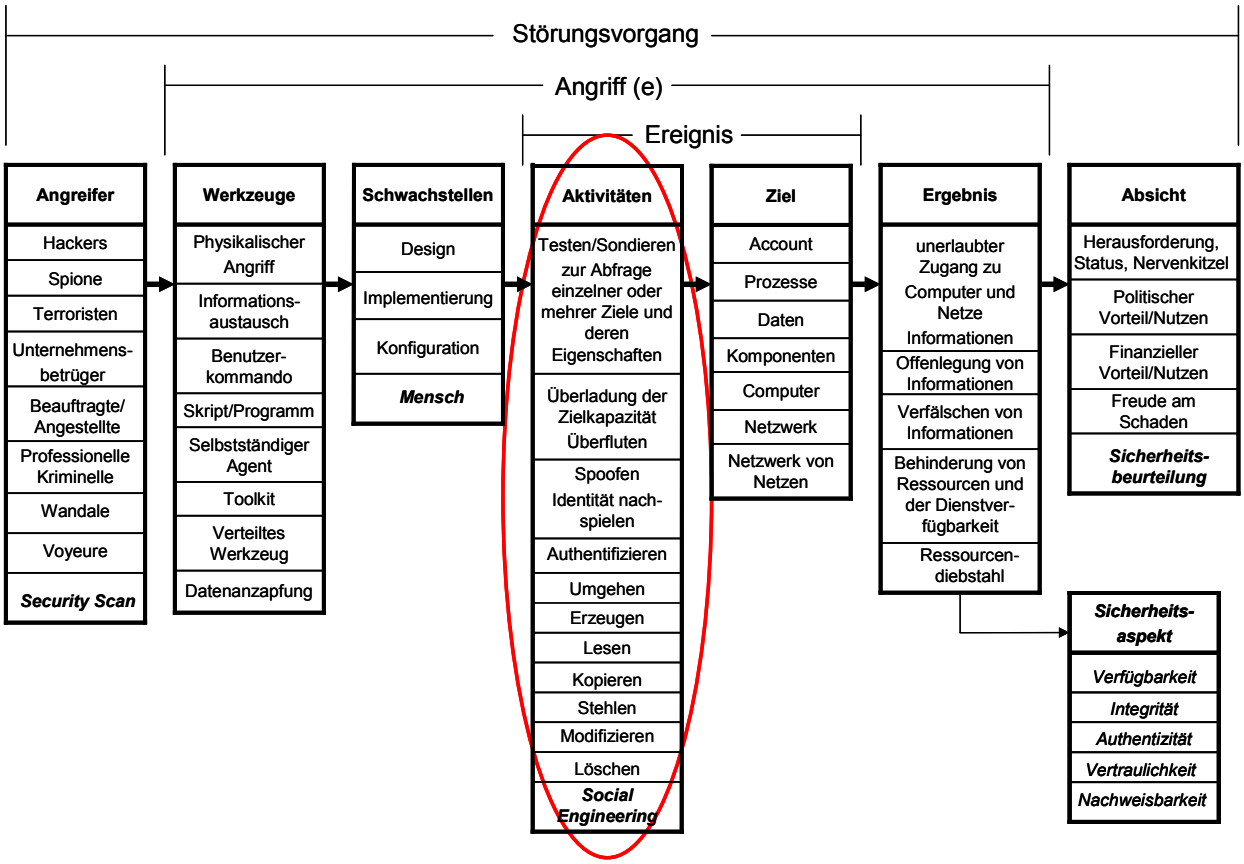

Abbildung 28: Computer und Netzwerk Störungsvorgang Taxonomie (CERT) mit Erweiterung Sicherheit.

In Bezug auf Angriffe und IT-Sicherheit muss zwischen autorisierten und unautorisierten Vorgehen/ Handlungen innerhalb eines Ereignisses unterschieden werden. Ein Benutzer loggt sich in einen Account ein. Dies ist grundsätzlich ein autorisiertes Vorgehen. Verfügt ein Benutzer nun über das Wissen der Account-Zugangsdaten eines ihm eigentlich nicht zugänglichen Accounts, so wird er vom System als der vermeintlich richtige Benutzer erkannt, ist es aber in Wirklichkeit nicht. Das System kann in diesem Fall nicht zwischen autorisierten und nicht autorisierten Benutzern unterscheiden. Dies ist dann ein unautorisiertes Vorgehen. Im Kontext der IT-Sicherheit muss klar getrennt werden können, wie an der Angreifer an die Information gelangt ist, da so der bzw. die betroffenen Sicherheitsaspekte bestimmt und zugeordnet werden können, das Schadensausmaß geschätzt werden kann und entsprechend brauchbare Mechanismen ausgewählt werden können.

Für Angriffe werden die Schwachstellen von Systemen genutzt. Laut CERT Taxonomie [HoLo03] werden drei Klassen von Schwachstellen unterschieden, wobei als vierte Klasse der Mensch hinzugefügt werden sollte (siehe Abbildung 28), wie dies bereits in [KLD07] ähnlich beschrieben ist.

1. Design
2. Implementierung
3. Konfiguration
4. Mensch

Ein Angreifer analysiert zunächst diese Schwachstellen beispielsweise mittels verschiedener (Sondierungs-)Tests. Aus Sicht eines Angreifers will dieser sich immer autorisiert erscheinenden Zugang verschaffen. Diesbezüglich steht das „Wie“ immer im Mittelpunkt eines Angreifers. Es gibt demnach verschiedene Aktivitäten, mit denen sich ein Angreifer autorisierten Zugang verschafft, um dadurch einer Identität oder einem System Schaden zuzufügen. Zu diesen Aktivitäten zählen: (Sondierungs-)Tests, Scannen, Überfluten, Überbrücken, Spoofen, Lesen, Kopieren, Stehlen, Modifizieren/ Neuerstellen und Löschen. In Tabelle 10 ist der Einfluss dieser Aktivitäten den IT-

Sicherheitsaspekten gegenübergestellt. Je nach Art der Aktivität können die einzelnen IT-Sicherheitsaspekte verletzt werden.

Tabelle 10: Potentielle Auswirkungen von Aktivitäten auf Sicherheitsaspekte.

	Verfügbarkeit	Integrität	Authentizität	Vertraulichkeit	Nachweisbarkeit
Testen Sondieren	X	X	X	X	X
Überfluten	X				
Authentifizieren			X	X	X
Überbrücken Umgehen			X		
Spoofen			X		
Lesen				X	
Kopieren	X		X	X	X
Stehlen	X	X	X	X	X
Modifizieren Erzeugen	X	X	X	X	X
Löschen	X	X	X		X

5.1.5 Angriffe

Angriffe dienen grundsätzlich dazu, unautorisierten Identitäten Zugang zu Information, Einrichtungen oder Ressourcen zu verschaffen. In Bezug auf digitale Kommunikationssysteme verfolgen Angreifer dabei unterschiedliche Strategien, die in den folgenden Abbildungen aufgeführt und verdeutlicht sind. Die Angriffe sind auf die Aktivitäten vgl. Abbildung 28 zurückzuführen.

Zunächst ist immer von einem *normalen Datenfluss* auszugehen. Ein Datenpaket wird von einer Informationsquelle zu einem Informationsziel gesendet und erreicht dieses unbeschadet. Die IT-Sicherheitsaspekte werden nicht verletzt. Bei der *Unterbrechung* erreicht das von einer Informationsquelle gesendete Informationspaket das Informationsziel nicht. Unterbrechungen können zum einen technischen Ursprungs sein, wie z.B. Übertragungsprobleme. Zum anderen können Unterbrechungen auch gezielt durch einen Angreifer inszeniert werden.

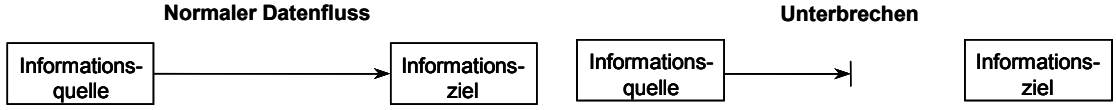

Abbildung 29: a) Normaler Datenfluss (links) und b) Unterbrechen (rechts).

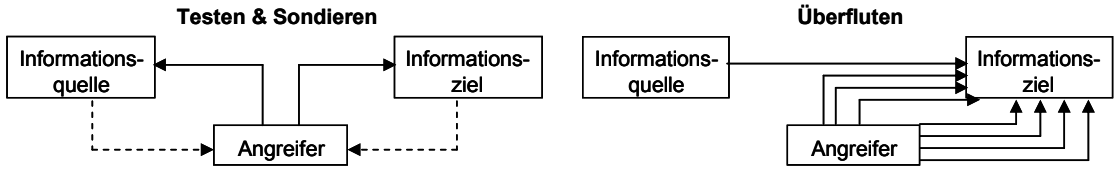

Abbildung 30: a) Testen und Sondieren (links) und b) Überfluten (rechts).

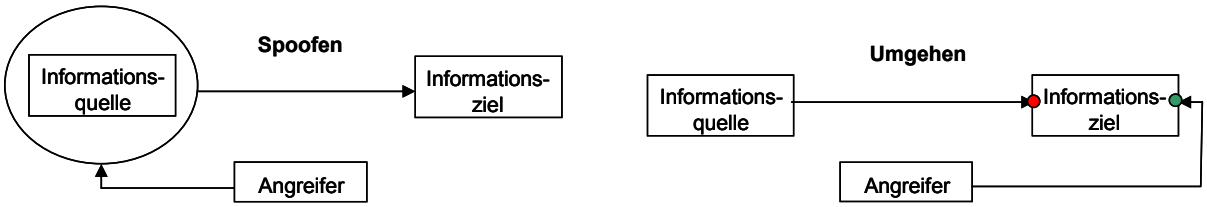

Abbildung 31: a) Spoofen (links) und b) Umgehen (rechts).

Spoofen ist das Annehmen bzw. Vortäuschen einer anderen Identität, um so unberechtigten Zugriff zu Daten zu bekommen und diese zu fälschen. Der Angreifer gibt vor, die Informationsquelle zu sein und erlangt so Zugriff auf das Informationsziel. *Spoofen* wird meist in Kombination mit anderen Angriffen benutzt, wie z.B. Lesen, Testen/ Sondieren, Social Engineering oder Authentifizieren. So wird zunächst Lesen, Testen/ Sondieren oder *Social Engineering* angewandt, um Information über die Informationsquelle zu erlangen. Diese Information, wie z.B. Passwörter wird dann benutzt, um zu *Spoofen* und so vom Informationsziel authentifiziert zu werden. *Spoofen* kann auch mit dem Ziel angewandt werden, Authentifizierungsverfahren zu untergraben, sie zu umgehen. Das *Umgehen* dient generell dazu, die Aktivierung von Prozessen zur Authentifizierung zu vermeiden. Es wird sich über einen anderen Weg Zugang zum Informationsziel verschafft.

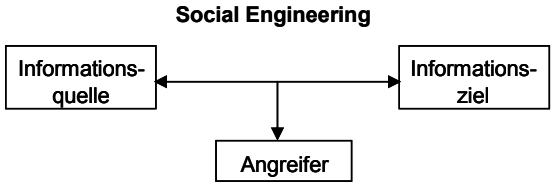

Abbildung 32: Social Engineering.

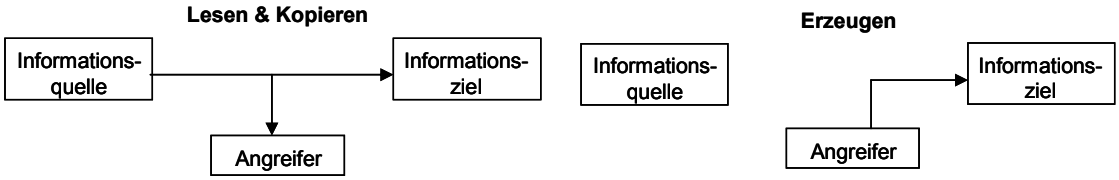

Abbildung 33: a) Lesen und Kopieren (links) und b) Erzeugen (rechts).

Beim *Erzeugen* sendet ein Angreifer Datenpakete an ein Informationsziel. *Spoofen* kann ebenfalls in die Kategorie Erzeugen fallen, nämlich dann wenn der Angreifer unter einer vorgegebenen bzw. gefälschten Identität neu erzeugte Datenpakete an das Informationsziel sendet. Das Versenden der Datenpakete kann nur auf die vorgegebene Identität zurückgeführt werden, nicht aber auf die des Angreifers.

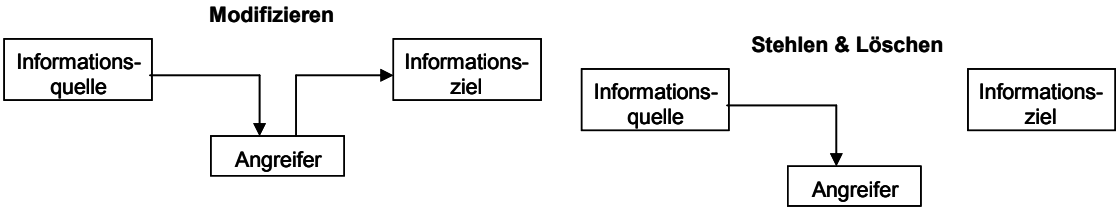

Abbildung 34: a) Modifizieren (links) und b) Stehlen und Löschen (rechts).

Beim *Stehlen und Löschen* werden Datenpakete vom Angreifer abgefangen bzw. vernichtet.

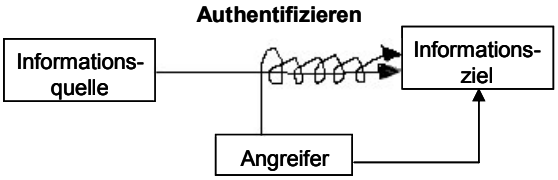

Abbildung 35: Authentifizieren.

Bekannte Angriffe

Tabelle 11 listet zu den bekanntesten und am meisten durchgeführten zählenden Angriffe (vgl. [LDKH06]).

Tabelle 11: Am bekanntesten und am meisten durchgeführte Angriffe, vgl. [LDKH06].

Angriff	Schwachstelle	Beispiele
Sniffing	Design, Implementierung, Konfigurierung, Mensch	Lesen
Man-in-the-Middle	Design, Implementierung, Konfigurierung, Mensch	Alle Aktivitäten, Lesen von Datenpaketen, Analysieren von Datenverkehr
Spoofing	Design, Implementierung, Konfigurierung, Mensch	Identität fälschen bzw. vortäuschen, Authentifizieren, meist in Kombination mit anderen Aktivitäten
Replay (Wiedergabe)	Design, Implementierung, Konfigurierung, Mensch; Ausnahmen in Abhängigkeit von der Topologie	Kombination Lesen (Sniffing) und Wiedereinspielen (Spoofing)
Denial of Service (Dienstverweigerung)	Design, Implementierung, Konfigurierung, Mensch	Überfluten
Malformed Packet	Implementierung, Design	„Ping of Death“
Malicious Code	Implementierung	Viren, Würmer, Trojanische Pferde, Zecken, …
Social Engineering	Mensch	Über Schulter schauen, Aushorchen, Erpressen, Mitlesen

5.1.6 Sicherheitsrichtlinien

In Bishop ist eine *Sicherheitsrichtlinie* (*Security Policy*) folgendermaßen definiert:

> *”Eine Sicherheitsrichtlinie ist eine Aussage, welche die Zustände eines Systems in eine Menge zulässiger oder sicherer Zustände und in eine Menge unzulässiger oder unsicherer Zustände einteilt. Der anzustrebende sichere Zustand ist dabei ein System, welches in einem zulässigen Zustand startet und keinen unzulässigen Zustand annehmen kann.“ [Bis03]*

Sicherheitsrichtlinien, auch *Security Policies* genannt, dienen zur Festlegung des erreichten bzw. des zu erreichenden Sicherheitsniveaus eines Systems, einschließlich aller darin enthaltenen Komponenten. Sicherheitsrichtlinien sind Pläne, welche die Ziele von Prozessen beschreiben, nicht aber die Prozesse selbst. Sie sind auch keine Guidelines, Standards oder Kontrollen. [Bar02] *IT-Sicherheitsrichtlinien* bezeichnen demnach Richtlinien zur Gewährleistung der notwendigen Sicherheit beim Einsatz von Informationstechnologien (IT). Dazu sind so genannte *IT-Sicherheitsstandards* definiert worden, welche die Erarbeitung und Umsetzung von *Sicherheitskonzepten* in den unterschiedlichen Bereichen betreffen. Als Grundlage dienen die vom Bundesamt für Sicherheit in der Informationstechnik (BSI) herausgegebenen Hilfsmittel, das IT-Grundschutzhandbuch (seit 2006 IT-Grundschutz-Kataloge) bzw. das IT-Sicherheitshandbuch. Sicherheitskonzepte beinhalten geeignete *Maßnahmen* (Mechanismen) zur Beseitigung bzw. zum Vorbeugen (Minimieren) in einer *Risikoanalyse* erkannter Gefahren und Risiken und berufen sich dabei auf bestimmte *Sicherheitsaspekte*. In einer Risikoanalyse wird ermittelt, welche technischen und organisatorischen Maßnahmen für den angestrebten *Schutzbedarf* (siehe BSI IT-Grundschutz-Kataloge) angemessen und erforderlich sind.

IT-Sicherheitsrichtlinien werden oft vernachlässigt. Als Hauptargumente für die Ignorierung von IT-Sicherheitsrichtlinien werden im Normalfall die finanzielle Situation verbunden mit dem zu hohen

Aufwand angeführt oder die so genannte im rechtlichen Sinne verstandene „billigende Inkaufnahme", in anderen Worten die „Es-wird-schon-nichts-passieren"-Einstellung. Das Vorhandensein und die Befolgung von IT-Sicherheitsrichtlinien ist aber die Voraussetzung eines einheitlichen Sicherheitsniveaus aller Komponenten eines Computersystems. Hierzu bedarf es eines so genannten *IT-Sicherheitsmanagements*.

5.1.7 IT-Sicherheitsmanagement

IT-Sicherheitsmanagement umfasst die im Folgenden aufgezählten Schwerpunkte. Hier ist anzumerken, dass diese Ausführungen etwas von denen in den BSI-Grundschutz-Katalogen abweichen. Im Sinne von Standardisierungen sei hier auf die BSI-Grundschutz-Kataloge verwiesen für eine detaillierte Darstellung der Methodik des IT-Grundschutzes wie der Initiierung des IT-Sicherheitsprozesses oder der Erstellung der IT-Sicherheitskonzeption im IT-Sicherheitsmanagement[25].

- Sensibilisierung für den Bedarf IT-Sicherheit
- Sicherheitspolitik
- Auswahl und Etablierung einer geeigneten Organisationsstruktur für IT-Sicherheit
- Risikoanalyse und Erstellung von Anforderungen in Sicherheitsrichtlinien (*Security Policies)*
- Auswahl und Umsetzung der geeigneten Maßnahmen, Anwendung geeigneter Mechanismen
- Integration von IT-Sicherheit in bestehende und laufende Prozesse

Zunächst muss grundsätzlich für den Bedarf an IT-Sicherheit sensibilisiert werden. Darauf aufbauend muss eine Sicherheitspolitik festgelegt werden, die alle folgenden Punkte enthält. Eine geeignete Organisationsstruktur für IT-Sicherheit muss ausgewählt und etabliert werden.

Verantwortlichkeiten müssen zugeteilt, Bedrohungspotential und Schadensausmaß müssen abgeschätzt werden (Risikoanalyse). Sicherheitsrichtlinien müssen festgelegt werden. Die konkrete Umsetzung muss geplant und organisiert werden, was die Auswahl und Umsetzung der geeigneten Maßnahmen (Mechanismen) beinhaltet. Am Ende wird IT-Sicherheit so in die bestehenden und laufenden Prozesse integriert.

5.1.8 Maßnahmen (Sicherheitsmechanismen)

In Bezug auf IT-Sicherheit ist zum Erreichen eines erwünschten Sicherheitsniveaus die Anwendung von Mechanismen bzw. Maßnahmen erforderlich. Diese *Maßnahmen* sind durch die Sicherheitsrichtlinien festgelegt und sind Bestandteil des Integrationsprozesses der IT-Sicherheit in die bestehenden und laufenden Prozesse. So wurden Mechanismen als Mittel zur Durchsetzung von Sicherheit definiert [Pfl03]. Die Leistungsfähigkeit solcher Maßnahmen kann durch Sicherheitsevaluierungen geprüft und zertifiziert werden. Eine Maßnahme ist also eine konkrete, an die Hardware, das Betriebssystem und die laufende Anwendungssoftware angepasste Umsetzung der Sicherheitsrichtlinie. Zu solchen Maßnahmen zählen unter anderen z.B.

- Hashverfahren (inhaltlich / kryptographisch)
- Kryptographie (wie z.B. digitale Signaturen)
- Biometrie
- Digitale Wasserzeichen
- Forensische Methoden
  - Computerforensik
  - Digitale Medienforensik

Im Bereich der *forensischen Methoden*, insbesondere der digitalen Medienforensik laufen derzeit viele Forschungsarbeiten und Entwicklungen, die auch für die digitale Langzeitarchivierung in Zukunft relevant sein können. Ansätze sind z.B. entwickelt und vorgestellt worden von [Far08a], [Far08b], [FrLG06], [JoFa07], [KMM+06], [LRF04], [LuFG05], [MAC+05], [MCA+05], [OeDi05], [OLD05],

[25] http://www.bsi.de/literat/bsi_standard/standard_1002.pdf

[PoFa04] oder [WaFa07]. Die hier aufgeführten Literaturangaben repräsentieren lediglich eine Auswahl und erheben keinen Anspruch auf Vollständigkeit.

Sicherheitsmechanismen dienen der (vgl. [Bis03]):

- **Prevention** Dem Vorbeugen von Verletzungen der IT-Sicherheitsaspekte
- **Detection** Dem Detektieren (wie Erkennen und Prüfen) von Verletzungen der IT-Sicherheitsaspekte
- **Recovery** Der Rekonstruktion von Verletzungen der IT-Sicherheitsaspekte

5.1.9 Common Criteria (CC)

Common Criteria (CC)[26] ist ein Standard, in dem gemeinsame Kriterien für die Prüfung und Bewertung der Sicherheit von Informationstechnik festgehalten sind. Sicherheitskonzepte und der Wechselbeziehungen einzelner Komponenten in Bezug auf Sicherheit können vgl. Abbildung 36 dargestellt werden. Sicherheit bedeutet Schutz von so genannten Assets vor Bedrohungen und Missbrauch.

Zu einem *Asset* zählen Information in Form von digitalen Objekten, Ressourcen, Gegenstände oder Güter, die einen Wert darstellen. Dies kann z.B. eine Datei sein, die etwas enthält, was in einem bestimmten Kontext verwendet werden soll, vorausgesetzt die Rechte zur Verwendung sind vorhanden. Sonst kann die Datei nicht benutzt werden, d.h. sie hat keinen Wert mehr und ist wertlos. Ein *Asset* muss verwertbar sein, d.h. Information über das *Asset* müssen vorhanden sein. Eine Datei ist demnach dann ein *Asset*, wenn Information über ihren Inhalt vorliegen (Dateiname, Metadaten). Eine Datei ist nicht verwertbar wenn nicht bekannt ist, was die Datei einhält.

Bedrohungen entstehen durch böswilliges, zielgerichtetes Zutun des Menschen mit der Absicht einem anderen Menschen oder einem System Schaden zuzufügen, Information zu verfälschen und Täuschungen herbeizuführen.

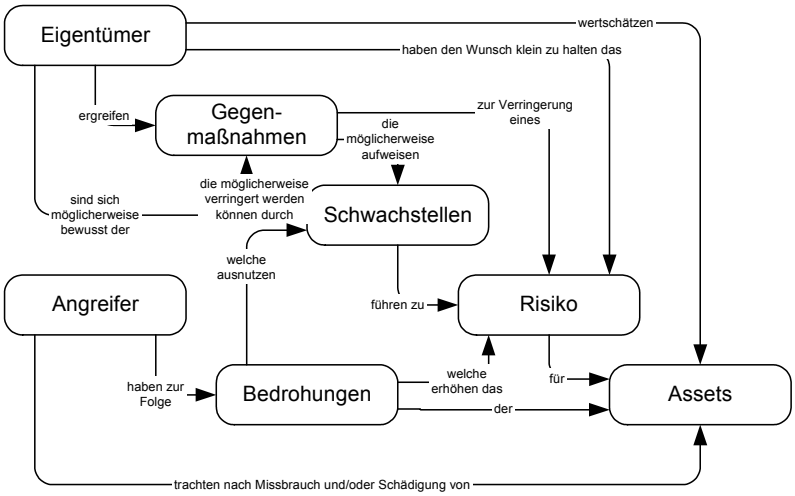

Abbildung 36: Sicherheitskonzepte und Wechselbeziehungen [CC05, Figure 2, S. 21] .

Die Verantwortung für das ursprüngliche Sicherheitsniveau liegt beim Eigentümer, indem er den *Assets* einen für ihn relevanten Wert zuspricht und so den *Schutzbedarf* des *Assets* festlegt. Der Eigentümer bestimmt also *was* und *inwieweit* das Was gesichert werden muss. Die Verantwortung bzgl. des *Wie* und des *Wann* es gesichert wird, also welche Mechanismen dazu bereitgestellt und herangezogen werden, wird, wie hier in Langzeitarchivierungssystemen, an das System, das Archiv selbst, weiterdelegiert. Es trägt dafür Sorge, die Integrität und Authentizität eines *Assets* zu bewahren und die Verfügbarkeit zu gewährleisten und hat die Aufgabe Sicherheitskonzepte dafür bereitzustellen.

[26] http://www.commoncriteriaportal.org/

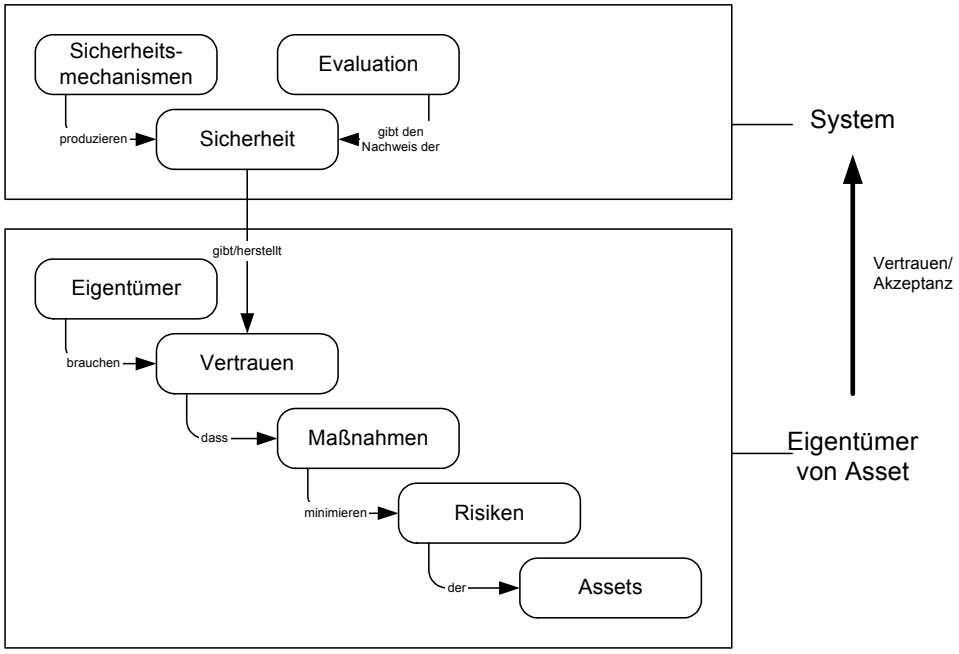

Abbildung 37: Vertrauenswürdige Systeme und Wechselbeziehungen, vgl. [CC05, Figure 3, S. 22].

Vgl. Abbildung 36 und Abbildung 37 wägt der Eigentümer eines *Assets* das Risiko ab, wie hoch die Bedrohung seines *Assets* sein kann. Auf dieser Basis entscheidet er, ob er dem System vertraut, es folglich akzeptiert. Um einem System zu vertrauen, muss der Eigentümer sicher sein, dass ausreichende und angemessene Maßnahmen getroffen werden, um Bedrohungen seines *Assets* zu senken bzw. zu minimieren. Da ein Eigentümer jedoch nicht alle beeinflussenden Faktoren beurteilen kann, ist eine Bewertung von potentiellen Maßnahmen notwendig. Sowohl die Maßnahmen als auch deren Bewertung ist vom System zu realisieren. Das Ergebnis solcher Bewertungen ist die Aussage, inwieweit den angebotenen Maßnahmen sicher und zuverlässig vertraut werden kann, dass sie in der Lage sind, das Risiko für potentielle Bedrohungen des Assets zu senken. Der Eigentümer bezieht diese Aussage nun in seine Entscheidung mit ein, ob er das Risiko eingeht sein *Asset* den verbleibenden Bedrohungen auszusetzen und folglich ob er, indem er sein *Asset* an das System übergibt, dem System vertraut und es akzeptiert. Der Kreislauf in Abbildung 37 schildert die Grundlage *vertrauenswürdiger Systeme*, wie sie zur Langzeitarchivierung benötigt werden. „Ein System ist dann vertrauenswürdig, wenn ein ausreichend glaubhafter und zuverlässiger Hinweis existiert, der zu glauben veranlasst, dass ein System bestimmten Anforderungen genügen wird.“ [Bis03]

Im Zusammenhang mit *Assets* ist das so genannte *Digital Asset Management* (DAM) bzw. *Media Asset Management* (MAM) zu erwähnen, Lösungen zur Aufbewahrung und Verwaltung von digitalen Objekten und Informationseinheiten bzw. Assets.

5.1.10 Dokumentation, Transparenz und Vertrauen in digitalen Langzeitarchiven

Die Aspekte Dokumentation und Transparenz dienen als Methoden zur Herstellung der Vertrauensbildung und tragen damit zur Entwicklung vertrauenswürdiger Systeme zur digitalen Langzeitarchivierung multimedialer Information bei.

Dokumentation

Im nestor-Kriterienkatalog ist Dokumentation wie folgt beschrieben:

> *„Die Ziele, die Konzeption und Spezifikation sowie die Implementierung des digitalen Langzeitarchivs sind angemessen zu dokumentieren. Anhand der Dokumentation kann der Entwicklungsstand intern und extern bewertet werden. Eine frühzeitige Bewertung kann auch dazu dienen, Fehler durch eine ungeeignete Implementierung zu vermeiden. Insbesondere erlaubt es eine angemessene Dokumentation aller Stufen, die Schlüssig-*

keit eines digitalen Langzeitarchivs umfassend zu bewerten. Auch alle Qualitäts- und Sicherheitsnormen fordern eine angemessene Dokumentation.“ [Nes06]

Transparenz

Transparenz in Bezug auf die vertrauenswürdige und abgesicherte Langzeitarchivierung digitaler Information ist im nestor-Kriterienkatalog folgendermaßen definiert:

„Transparenz wird realisiert durch die Veröffentlichung geeigneter Teile der Dokumentation. Transparenz nach außen gegenüber Nutzern und Partnern ermöglicht diesen, selbst den Grad an Vertrauenswürdigkeit festzustellen. Transparenz gegenüber Produzenten und Lieferanten bietet diesen die Möglichkeit zu bewerten, wem sie ihre digitalen Objekte anvertrauen. Die Transparenz nach innen dokumentiert gegenüber den Betreibern, den Trägern, dem Management sowie den Mitarbeitern die angemessene Qualität des digitalen Langzeitarchivs und sichert die Nachvollziehbarkeit der Maßnahmen. Bei denjenigen Teilen der Dokumentation, die für die breite Öffentlichkeit nicht geeignet sind (z.B. Firmengeheimnisse, Information mit Sicherheitsbezug), kann die Transparenz auf einen ausgewählten Kreis (z.B. zertifizierende Stelle) beschränkt werden. Durch das Prinzip der Transparenz wird Vertrauen aufgebaut, da es die unmittelbare Bewertung der Qualität eines digitalen Langzeitarchivs durch Interessierte zulässt.“ [Nes06]

Dokumentation und Transparenz sind im digitalen Langzeitarchiv zusammenhängend zu betrachten. Vertrauen wird durch Transparenz hergestellt. Zur Realisierung der Transparenz ist eine Dokumentation bzgl. verschiedener Aspekte des digitalen Langzeitarchivs (Ziele, Konzeption, Spezifikation, Implementierung, insbesondere Qualitäts- und Sicherheitsnormen) notwendig. Es bedarf hier einer genauen Abschätzung inwieweit dokumentiert werden muss bzw. soll unter der Berücksichtigung der möglichen Angreifbarkeit und was das Archiv an Dokumentation überhaupt leisten kann.

Vertrauen

Vertrauenswürdige Systeme werden dadurch erreicht, dass Vertrauen hergestellt wird und daraus resultierend die Akzeptanz des Systems seitens des Benutzers. Dabei wird oft vergessen, dass das Vertrauen aber auch beibehalten werden muss und dies ein dynamischer Prozess ist. Vertrauen seitens des Benutzers ist kein fester Zustand, sondern ändert sich ständig über die Zeit. Inwiefern Vertrauen mit IT-Sicherheit verbunden ist, kann in der Literatur nachgelesen werden [OeDi06].

5.1.11 Sicherheitsaspekte im Kontext der Langzeitarchivierung

Dieser Abschnitt der Expertise stellt Einsatzmöglichkeiten von Sicherheitstechnologien für eine Gewährleistung einer vertrauenswürdigen und abgesicherten Langzeitarchivierung vor und evaluiert deren Eignung. Hierzu werden zunächst die Sicherheitsaspekte Verfügbarkeit, Integrität, Authentizität, Vertraulichkeit und Nachweisbarkeit im Kontext der digitalen Langzeitarchivierung beschrieben.

Die Evaluation orientiert sich an den Sicherheitsaspekte. Zu jedem einzelnen dieser Sicherheitsaspekte führt sie dessen Ausprägungen als Anforderungen an eine vertrauenswürdige und abgesicherte digitale Langzeitarchivierung auf. Diese Anforderungen repräsentieren den idealen Soll-Zustand. Für jede Anforderung werden Sicherheitsmechanismen benannt, welche grundsätzlich dazu geeignet sind, zur Erfüllung der Anforderung beizutragen.

Für die Reflektion der Sicherheitstechnologien in Bezug auf die organisatorischen und technischen Rahmenbedingungen innerhalb der betrachteten Szenarien werden die Anforderungen mit den grundsätzlich zu ihrer Erfüllung geeigneten Sicherheitstechnologien mit der Erhebung des Ist-Zustandes vervollständigt.

Verfügbarkeit

Verfügbarkeit in digitalen Langzeitarchiven bedeutet, dass archivierte Daten über einen langen Zeitraum für eine bestimmte Benutzergruppe jederzeit zugänglich sein sollen. Das schließt die Ver-

fügbarkeit des Speichermediums, der Systemkomponenten sowie des Darstellungsprogramms ein. Verfügbarkeit im Zusammenhang digitaler Langzeitarchive sollte demnach unterteilt werden in:

- Verfügbarkeit der Systemressourcen einschließlich der Speichermedien und Datenbanken
- Verfügbarkeit der Darstellungsanwendungen
- Verfügbarkeit des digitalen Archivobjektes

Verfügbarkeit bedeutet im Kern die *Wiederauffindbarkeit* von Objekten im Archiv, was sowohl den physischen Speicherort, die logische Interpretation als auch die konzeptuelle Darstellung betrifft. Zur Verfügbarkeit in digitalen Langzeitarchiven müssen Mechanismen vorhanden sein, welche sicherstellen, dass Objekte bei Bedarf im Speicher auch in Zukunft wieder gefunden werden können, unabhängig davon, ob der Speicher sich verändert hat, denn in digitalen Langzeitarchiven muss davon ausgegangen werden, dass ein Speichermedium aufgrund seiner Alterung im Archiv innerhalb bestandserhaltender Maßnahmen ausgetauscht wird. Die Verfügbarkeit in digitalen Langzeitarchiven schließt darüber hinaus die Aufbereitung des digitalen Objektes je nach Zugriffsrechten und Darstellungsanwendung und –Medium ein.

Integrität

Die Bedeutung der Integrität im Zusammenhang mit der Langzeitarchivierung gestaltet sich wie folgt: Eine wesentliche Anforderung an die Langzeitarchivierung digitaler Information besteht darin, dass archivierte Daten fortdauernd über einen langen Zeitraum vollständig und unverändert vorliegen sollen und dementsprechend integer erhalten werden müssen. Die Bedrohung der Integrität ist im digitalen Langzeitarchiv bei allen Prozessen gegenwärtig, sowohl beim *Ingest* oder *Access* als auch bei Verfahren der Bestandserhaltung innerhalb des Archivs. Die digitalen Archivobjekte werden kopiert, transformiert und migriert. Sie werden auf verschiedenen physischen Speichermedien abgelegt, müssen zur Weiterverarbeitung von verschiedenen Anwendungen interpretiert werden können und je nach Medientyp für ein Wahrnehmungssystem, wie das der Menschen (Hören, Sehen) dargestellt werden.

Folgt man der Darstellung des digitalen Objektes von Thibodeau [Thi02], bezieht sich die Integrität demnach auf alle Ebenen eines digitalen Objektes. Die Sicherung der Integrität auf *physischer Ebene* betrifft die Erhaltung des Speichermediums, um die Daten unverändert, vollständig und fehlerfrei auch in der Zukunft lesen zu können. Die Sicherung der Integrität auf *logischer Ebene* betrifft das logische Datenmodell, Datenschema und Datenformat, in dem Information individuell gespeichert ist, insbesondere deren vollständige Erhaltung, um von Anwendungen korrekt interpretierbar zu bleiben. Die Sicherung der Integrität auf *konzeptueller Ebene* betrifft die Darstellung der Information und die Komponenten, die dafür benötigt werden, die Information korrekt und gleich bleibend zu interpretieren, um den eigentlichen Inhalt darzustellen.

Authentizität

Die Authentizität ist im digitalen Langzeitarchiv vor allem bei der Einstellung/ Aufnahme von neuen Archivobjekten (*Ingest*) bedroht aber auch während der Weiterverarbeitung einschließlich der Auslieferung (*Access*) sowie bei der Anwendung von Verfahren der Bestandserhaltung innerhalb des Archivs.

Authentizität in digitalen Langzeitarchiven besagt, dass die Originalität und der Ursprung der archivierten Objekte erhalten bleiben, was bedeutet, dass ein späterer Zugriff tatsächlich auf dem gleichen originären Archivobjekt geschieht, welches ursprünglich ins Archiv aufgenommen wurde. Die Sicherung der Authentizität im digitalen Langzeitarchiv umfasst zum einen Maßnahmen, die Veränderungen bzw. Manipulationen nach der Aufnahme eines Objektes ins Archiv verhindern und zum anderen Maßnahmen die nachweisen, dass keine Veränderungen bzw. Manipulationen nach der Aufnahme eines Objektes ins Archiv stattgefunden haben. Nachträgliche Veränderungen an archivierten Objekten sind zur Authentizitätssicherung durch Protokolle und eine Versionsverwaltung nachvollziehbar nachzuweisen. Darüber hinaus beinhaltet Authentizität in digitalen Langzeitarchiven

die nachweisliche Überprüfbarkeit des Autors und des Einstellungszeitpunktes im Zusammenhang der Einhaltung rechtlicher Vorschriften.

Vertraulichkeit

Vertraulichkeit in digitalen Langzeitarchiven heißt, dass ein Zugriff auf die archivierten Daten nur von vorher autorisierten Personen erlaubt ist. Maßnahmen zur Sicherung der Vertraulichkeit müssen davor schützen, dass kein unbefugter Zugriff auf die Archivobjekte geschehen kann, was bedeutet, dass Zugriffsaktivitäten voneinander abgeschirmt werden müssen.

Die allgemeinen Zugriffsrechte sowie die systembedingten technischen Zugriffsrechte bedürfen einer eindeutigen Regelung. Ebenso die Zugriffsbeschränkungen auf das Archivsystem und die Zugriffsrechte im Umgang mit digitalen Archivobjekten. Die Zugriffsrechte müssen eindeutig festgelegt sein. In der Regel wird bezüglich des Zugriffs auf das Archivsystem der Administration die meisten Rechte zugesprochen während im Umgang mit den digitalen Objekten der Zugriff je nach Objekt variiert. Zur Vorbeugung von Verletzungen der Vertraulichkeit sollte die Gestaltung der Zugriffsrechte die Unterscheidung von erlaubten und unerlaubten Handlungen einbeziehen. Darüber hinaus beinhaltet die Vertraulichkeit in digitalen Langzeitarchiven den Schutz der Urheberrechte und die Regelung der Nutzungsrechte. Ein digitales Langzeitarchiv sollte demnach die unerlaubte Vervielfältigung von urheberrechtlich geschützten Archivobjekten unterbinden können.

Nachweisbarkeit (Nicht-Abstreitbarkeit)

Die Nachweisbarkeit bzw. Nicht-Abstreitbarkeit hat im Zusammenhang mit digitalen Langzeitarchiven folgende Bedeutung: Nachweisbarkeit (Nicht-Abstreitbarkeit) beinhaltet, dass der Ursprung von Information und deren Urheber unabstreitbar und verbindlich nachgewiesen werden kann. Das Langzeitarchiv muss sicherstellen, dass zu einer Anfrage das entsprechende Archivobjekt nachweislich ausgeliefert wurde. Darüber hinaus beinhalten archivierte Objekte teilweise den Nachweis, dass ein bestimmtes Ereignis zu einem bestimmten Zeitpunkt stattgefunden hat, wie etwa ein Vertrag, Bild- oder Tonaufnahmen usw., zustande gekommen ist. Demnach muss das Archivobjekt in seiner ursprünglichen Gültigkeit und Form erhalten bleiben. Dies ist nicht nur eine Anforderung an bestimmte Langzeitarchivierungssysteme sondern ebenso eine Herausforderung, denn es müssen geeignete Mechanismen eingesetzt werden, die eine solche Gültigkeit über einen langen Zeitraum sicherstellen und die selbst für diesen Zeitraum gültig und verfügbar sind. Die Nachweisbarkeit vereint die zuvor genannten Sicherheitsaspekte Verfügbarkeit, Vertraulichkeit, Integrität und Authentizität.

Sicherheitsmechanismen

In diesem Abschnitt werden die Sicherheitsmechanismen erhoben, die für eine vertrauenswürdige und abgesicherte Langzeitarchivierung angewandt werden können und sollten. In den Tabellen werden die Sicherheitsmechanismen den spezifischen Anforderungen, die sich aus den Sicherheitsaspekten ergeben, zugeordnet. Darüber hinaus wird der Ist-Zustand benannt, auf dem aufbauend der zur Überbrückung der Differenz zu den Soll-Anforderungen notwendige Handlungsbedarf in Kapitel 6 beschrieben wird. Die Sicherheitsmechanismen werden vorher nicht explizit kategorisiert und stattdessen direkt den Anforderungen gegenübergestellt. Als Sicherheitsmechanismen sind nicht nur explizit Mechanismen der IT-Sicherheit aufgeführt wie z.B. Kryptographie, digitale Signaturen, digitale Wasserzeichen, forensische Methoden, Biometrie, VPN, Firewalls, verteilte Speicherlösungen oder Grid. Zu den Sicherheitsmechanismen in digitalen Langzeitarchivierungssystemen zählen auch generelle Erhaltungsmaßnahmen, wie Migration oder Emulation, sowie Metadaten, Identifier, etc.

Sicherheitsmechanismen werden generell unterschieden in Mechanismen, die Medienbrüche/n

a) Vorbeugen (*Prevention*)
b) Erkennen und Prüfen (*Detection*)
c) Rekonstruieren (*Recovery*)

Zusätzlich gibt es Sicherheitsmechanismen, die Medienbrüche verursachen. Ein Sicherheitsmechanismus, der in einem bestimmten Kontext und mit einem bestimmten Ziel eingesetzt wird, kann in einem anderen Zusammenhang eine andere Wirkung haben. So kann ein Mechanismus beispielsweise gleichzeitig Prüfen und Vorbeugen.

Weiterhin muss zwischen Medienbrüchen, Modellbrüchen, Schemabrüchen und Formatbrüchen unterscheiden werden, deren Auswirkungen im Archiv unterschiedlich stark sind. Bei Medienbrüchen sind Sicherheitsmechanismen zur Prüfung und Rekonstruktion nur schwer einsetzbar. Auch mit Referenzmaterial gestaltet sich dies sehr schwer, während sich Formatbrüche mit Hilfe von Referenzmaterial prüfen und rekonstruieren lassen.

5.1.12 Erhebung der verwendbaren Sicherheitsmechanismen in Bezug auf die Anforderungen

Ausgehend von den Sicherheitsaspekten werden in den folgenden Tabellen die Sicherheitsmechanismen mit ihrer Zuordnung zu den Anforderungen erhoben. Dadurch wird eine übersichtliche und strukturierte Herangehensweise erlangt, um den Ist-Zustand der Integration von Sicherheitsmechanismen in derzeitigen digitalen Langzeitarchiven aufzuzeigen. In den Tabellen ist eine Auswahl von Sicherheitsmechanismen aufgeführt. Dieses tabellarische Gegenüberstellungsschema erhebt keinen Anspruch auf Vollständigkeit und ist offen für Anpassungen und Erweiterungen, um spezifischen Archiven und Anwendungsfeldern Rechnung zu tragen. Die Erfassung des Ist-Zustands erfolgt im Abschnitt 5.2. An dieser Stelle wird das Schema hergeleitet und die Bedeutung der Kriterien erläutert.

Tabelle 12: Verfügbarkeit (exemplarisch).

Anforderung (Soll)		**Sicherheitsmechanismus**	**Zustand (Ist)**
Digitales Archivobjekt	Wiederauffindbarkeit	Identifikatoren	
		Index	
		Suchfunktion	
	Zugriff Versionen	Versionsverwaltung	
		Historie	
	Vorhandensein	Antivirenprogramm	
	Interpretierbarkeit	Darstellungsanwendung	
Systemressourcen (Komponenten)	Hardware-Netzwerkaufbau	Grid	
		Verteilte Speicherlösungen	
	Software - Netzwerkverbindungen	Zugänge	
		Clients	
		Firewall	
		VPN	
Darstellungsanwendungen		Metadaten	
		Emulationen	

Tabelle 12 beinhaltet die Sicherheitsmechanismen, die den speziellen Soll-Anforderungen für den Sicherheitsaspekt Verfügbarkeit zugeordnet sind. Diese Anforderungen sind unterteilt in die Bereiche:

- Digitales Archivobjekt
- Systemressourcen (Komponenten)
- Darstellungsanwendung

Um ein archiviertes digitales Objekt zu adressieren, z.B. zum Zeitpunkt des erneuten Zugriffs, sollten so genannte Identifikatoren angewandt werden. Ebenso sollte ein Index sowie Suchfunktionen zur Verfügung stehen. Generell muss als Sicherheitsmechanismus eine Antivirensoftware im digitalen

Langzeitarchiv zum Einsatz kommen, um die langfristige Verfügbarkeit von Archivobjekten sicherzustellen. Ein Schadensszenario ist hier, dass digitale Objekte mit Schadprogrammen wie Viren oder Trojanischen Pferden behaftet ins Archiv eingespeist werden. Sowohl das mit dem Schadprogramm behaftete Objekt als auch andere Objekte werden nach einer bestimmten Zeit im Archiv zerstört. Mittels Antivirensoftware kann zum Großteil sichergestellt werden, dass Archivobjekte nicht durch Schadsoftware gelöscht werden.

Die Sicherheitsmechanismen bzgl. der Systemressourcen (Komponenten) betreffen zum einen die Hardware und den Netzwerkaufbau und zum anderen die Software und die Netzwerkverbindungen. Die Frage bzgl. des Netzwerkaufbaus ist hier, inwieweit Grid und verteilte Speicherlösungen dazu beitragen, abgesicherte Langzeitarchivierungssysteme zu realisieren. Bezüglich der Netzwerkverbindungen sollten Sicherheitsmechanismen angewandt werden, die eine abgesicherte Übertragung der Information ermöglichen wie z.B. VPN Clients, spezielle Zugänge oder Firewalls.

Des Weiteren ist es Aufgabe eines digitalen Langzeitarchivs dafür zu sorgen, dass die benötigte Darstellungsanwendung zur Interpretation des digitalen Archivobjektes verfügbar ist. Dazu werden beispielsweise Metadaten angewandt, die das Format der Speicherung dokumentieren und/ oder Information bereitstellen, die mindestens eine Transformation beschreiben, die das Objekt in ein zum gegenwärtigen Zeitpunkt maschinen- und menschenlesbares Format umwandelt. Ebenfalls können Emulationen eingesetzt werden, um das Darstellungsprogramm verfügbar zu machen.

Tabelle 13: Integrität (exemplarisch).

Anforderung (Soll)			**Sicherheitsmechanismus**	**Zustand (Ist)**
Digitales Archivobjekt	Physische Ebene		Migration	
			Umkopieren	
			Spiegelung	
			Backup	
			Verschlüsselung - Kryptographie	
			Hashfunktionen	
			Antivirenprogramm	
	Logische Ebene		Migration	
			Transformation	
			Digitales Wasserzeichen	
			Verschlüsselung – Kryptographie	
			Hashfunktionen	
			Antivirenprogramm	
	Konzeptuelle Ebene	Unveränderter Inhalt	Statistische Auswertungen	
			Objektive Auswertungen	
			Subjektive Auswertungen - Wahrnehmungstests	
			Antivirenprogramm	
Systemressourcen (Komponenten und Netzwerk)	Hardware	Unversehrtheit	Zugangsbeschränkung	
			„Käfige“	
			Trusted Computing (TCP), NTCB	
	Software	Unversehrtheit	IDS	

		Hashfunktionen – Fingerabdrücke a) Online (*Code*) b) Offline (*Dump*)
		Forensik
		Antivirenprogramm
Darstellungs-anwendungen	Unversehrtheit	Hashfunktionen – Fingerabdrücke a) Online (*Code*) b) Offline (*Dump*)
		Forensik
		Antivirenprogramm

Tabelle 13 beinhaltet die Sicherheitsmechanismen, die den speziellen Soll-Anforderungen für den Sicherheitsaspekt Integrität zugeordnet sind. Diese Anforderungen sind wie bereits zuvor unterteilt in die Bereiche:

- Digitales Archivobjekt
- Systemressourcen (Komponenten)
- Darstellungsanwendung

Für die Erfüllung der Sollanforderungen des Sicherheitsaspektes Integrität kommen Antivirenprogramme in allen Bereichen zum Einsatz.

Die Integrität eines digitalen Archivobjektes muss entsprechend Thibodeau auf drei verschiedenen Ebenen sichergestellt werden, der physischen Ebene, der logischen Ebene und der konzeptuellen Ebene. Die Migration ist dabei als Sicherheitsmechanismus zur Erhaltung der Integrität zu werten und kann sich einerseits ausschließlich auf die physische Ebene beschränken, kann andererseits aber auch die physische und logische und ggf. auch die konzeptuelle Ebene betreffen.

Eine Migration in Form von Umkopieren geschieht nur auf physischer Ebene während eine Migration in Form von Transformationen von der logischen Ebene ausgehend auch Auswirkungen auf die physische Ebene und ggf. die konzeptuelle Ebene hat. So kommt es beispielsweise bei einer verlustbehafteten Formatkompression zu einem Verlust von Information und somit der Verletzung der Integrität auf konzeptueller Ebene.

Digitale Wasserzeichen dienen der Prüfung der Integrität auf logischer Ebene. Die Anwendung digitaler Wasserzeichen lässt Veränderungen eines digitalen Objektes sehr gut nachvollziehen, kann dem jedoch nicht vorbeugen. Digitale Wasserzeichen haben den Nachteil, dass sie das Signal beeinflussen und somit den Inhalt verändern. Dies hat zur Folge, dass die Originalität bzw. Authentizität nicht mehr gewährleistet ist und leichte Qualitätseinbußen zu verzeichnen sind.

Hashfunktionen zur Überprüfung von Integritätsverletzungen hingegen werden bereits heute in Archiven digitaler Information angewandt, da sie leicht einsetzbar sind. Hashfunktionen bilden so genannte Checksummen und können so Veränderungen bestimmen. Mittels Hashfunktionen kann jegliche Art von Software überprüft werden. Dazu zählen Formate ebenso wie Darstellungsanwendungen. Eine Checksummenbildung mittels Hashfunktionen kann offline und online geschehen. Offline wird eine Checksumme über die ausführbare Programmdatei im Langzeitspeicher, online wird eine Checksumme über die ausführbare Programmdatei im Laufzeitspeicher gebildet.

Verschlüsselungstechniken der Kryptographie dienen der Vorbeugung von Verletzungen der Integrität, können aber bei Verlust des Schlüssels dasselbe auch verursachen.

Statistische, objektive und subjektive Auswertungen werden als Sicherheitsmechanismen aufgeführt für Messungen und Prüfungen von Integritätsverletzungen, insbesondere für die Messung von Qualitätsunterschieden. Hierbei ist in der Regel Referenzmaterial des Originals nötig, um digitale

Objekte auf Veränderungen zu prüfen und Qualitätsunterschiede festzustellen. Statistische Messgrößen sind beispielsweise Entropie und Informationsgehalt. Für die objektiven Auswertungen wird versucht, das menschliche Gehör maschinell nachzubilden während für subjektive Auswertungen Testpersonen Qualitätsunterschiede benennen müssen.

Methoden der Medienforensik ermöglichen ebenfalls eine Überprüfung von Verletzungen der Integrität und bieten darüber hinaus Ansätze zur Rekonstruktion und Wiederherstellung.

Bezüglich der Anforderung der Unversehrtheit (Integrität) der Systemressourcen (Komponenten und Netzwerk) wurde in Tabelle 13 *Trusted Computing* (TC) und NTCB als Sicherheitsmechanismen zugeordnet. Unter *Trusted Computing* werden Technologien zur Schaffung einer vertrauenswürdigen Plattform zusammengefasst. Im Zusammenhang mit *Trusted Computing* existiert eine so genannte *Trusted Computing Platform Alliance* (TCPA), die im Oktober 1999 gegründet wurde. Im Juli 2001 wurde der TCPA-Standard 1.1 vorgestellt und im Mai 2002 der TCPA-Standard 1.1b. Die genannte *Trusted Computing Group* (TCG)[27] gilt als offizieller Nachfolger der TCPA und hat als unmittelbares Ziel, den TCPA-Standard 1.2 zu entwickeln [Mei03]. Die TCG ist eine Industrievereinigung zur Steigerung der Vertrauenswürdigkeit von Rechnern und IT-Systemen durch Entwicklung von Standardspezifikationen. Die TCP sieht es als ihre Aufgabe an, eine Sicherheit zu entwickeln für Bedrohungen aus offenen Netzen, wie dem Internet, was auch im Kontext der digitalen Langzeitarchivierung von Bedeutung ist. Die TCPA/TCG Spezifikation soll in erster Linie das Vertrauen der Nutzer in netzwerkbasierte IT-Systeme herstellen und vertrauenswürdige Systeme anbieten.

Die Arbeit der TCG beinhaltet folgende Lösungen:

- Hardware-Chip gegen Bedrohungen aus dem Internet
- Schutz von Software vor Software

Dabei werden die folgenden konkreten Ziele verfolgt:

1. Beglaubigung (Attestation) der Systemintegrität gegenüber Dritten
2. Sicherung von Daten gegen Angriffe
3. Sicherung von kryptographischen Schlüsseln gegen Angriffe
4. Bereitstellung kryptographischer Funktionen: Hash, Zufallsgenerator

Bisher ist dabei kein Schutz vor Hardware-Angriffen enthalten. Die Bauweise und die Funktionsweise als *Trusted Platform Module* mit *Trusted platform Support Service*: TSS können auf der Website der TCG Group[28] nachgelesen werden.

Innerhalb der TCG gibt es eine so genannte *Storage Work Group*, die sich damit beschäftigt, Standardspezifikationen für die Sicherheit in Bezug auf geeignete Speichersysteme zu erarbeiten. Ihr Ziel ist die Entwicklung einer flexiblen Vertrauensarchitektur, die für verschiedene Speicherumgebungen und Anforderungen angewendet werden kann. Mit der Spezifikation soll erreicht werden, dass Plattformbasierte Anwendungen den Vorteil von Vertrauen und Sicherheit nutzen können, der durch Dienste eines vertrauenswürdigen Speichersystems angeboten werden kann. Beispiele dafür sind:

- Kryptographie
- Public-Key-Kryptographie und digitale Signatur
- Hashfunktionen
- Random Number Generation (RNG)
- Secure Storage

[27] https://www.trustedcomputinggroup.org

[28] https://www.trustedcomputinggroup.org

Die Spezifikationen der TCG beziehen sich auf die Sicherheit von Systemkomponenten, während die so genannte *Network Trusted Computing Base* (NTCB), bestehend aus Hardware und Software, bemüht ist, eine Netzwerksicherheit herzustellen[29].

Tabelle 14: Authentizität (exemplarisch).

Anforderung (Soll)		Sicherheitsmechanismus	Zustand (Ist)
Vorbeugen Änderungen – Erhaltung Original		Emulation	
		Antivirenprogramm	
		Firewall	
Nachweis unverändertes Original		Einfache Signatur	
		Fortgeschrittene Signatur	
		Qualifizierte Signatur	
		Hashfunktionen	
		Digitale Medienforensik	
Nachweis Änderungen		Protokolle	
		Versionsverwaltung	
		Digitale Wasserzeichen	
		Digitale Medienforensik	
Nachweis Autor		Digitale Signatur	
		Digitales Wasserzeichen	
Nachweis Quelle		Digitale Medienforensik	
Nachweis Archivzeitpunkt		Zeitstempel	
		Digitale Signatur	
		Digitale Medienforensik	
		Digitale Computerforensik	
Nachweis Archiv		Registrierung	
		Protokolle	
Authentifizierung	Administration	Kennwörter	
	Produzent	Schlüssel	
	Konsument	Biometrie	
	System-komponenten (Server)	IP, Protokolle	
	Angreifer	IDS	
		Protokolle	
		Computerforensik	

Tabelle 14 beinhaltet die Sicherheitsmechanismen, die den speziellen Soll-Anforderungen für den Sicherheitsaspekt Authentizität zugeordnet sind. Diese Anforderungen betreffen zunächst die Erhaltung des Originals einschließlich der Vorbeugung von Änderungen des originär ins Archiv aufgenommenen Objekts. Die Emulation wird zuverlässig als Sicherheitsmechanismus für die Erhaltung und Benutzung originaler digitaler Archivobjekte eingesetzt. Darüber hinaus sollten Antiviren-

[29] Network Trusted Computing Base
http://publib.boulder.ibm.com/infocenter/systems/index.jsp?topic=/com.ibm.aix.security/doc/security/tcpip_security_ntcb.htm

programme sowie Firewalls zum Einsatz kommen, denn sie sichern ein unverändertes Vorhandensein des Originals im Archiv. Der geforderte Nachweis über unverändertes Original kann mittels digitaler Signaturen, Hashfunktionen oder Methoden der digitalen Forensik erbracht werden. Für die Erhaltung der Authentizität im digitalen Archiv müssen Änderungen nachvollziehbar festgehalten sein. Dazu sollten beispielsweise Sicherheitsmechanismen wie Protokolle, Versionsverwaltungen, Methoden der digitalen Medienforensik oder digitale Wasserzeichen angewandt werden.

Der Nachweis des Autors kann durch digitale Signaturen oder digitale Wasserzeichentechnologien erbracht werden, während für den Nachweis der Quelle (Sensor/ Aufnahme- bzw. Digitalisierungsgerät) Methoden der digitalen Medienforensik angewandt werden sollten. Für die Sicherung der Authentizität sollte darüber hinaus der Archivzeitpunkt nachgewiesen werden können, wozu Zeitstempel, digitale Signaturen sowie Methoden der digitalen Computer- und Medienforensik eingesetzt werden sollten. In Bezug auf Zeitstempel ist eine Entscheidung notwendig, ob die einfache Zeitangabe ausreichend ist, oder ob Mechanismen zur Sicherstellung der Unveränderbarkeit der Zeitangabe (ist die Zeitangabe die originale und richtige) zusätzlich eingesetzt werden müssen. Weiterhin muss zur Sicherstellung der Authentizität nachgewiesen werden können, dass ein Archivobjekt in das richtige, dafür vorgesehene Archiv eingeliefert wurde. Dies kann beispielsweise über Registrierungen und Protokolle realisiert werden.

Im Zusammenhang mit der Authentizität ist die Authentifizierung als ein weiterer Bereich der Soll-Anforderungen an ein abgesichertes Langzeitarchiv digitaler Information aufzuführen. Authentifiziert werden müssen sowohl Administration, Produzent, Konsument als auch Systemressourcen (Komponenten). Dazu sollten Sicherheitsmechanismen wie Kennwörter und Schlüssel aber auch Biometrie zum Einsatz kommen. Die Authentifizierung von Systemressourcen (Komponenten) sollte über Protokolle und ihre IP- Adressen oder MAC-Adressen erfolgen.

Mit der Authentifizierung von Angreifern ist das Feststellen von Angriffen die Rückverfolgung bis zur Identifizierung des Angreifers gemeint. Hier sollten IDS-Systeme, Protokolle und Verfahren der Computerforensik Anwendung finden. IDS sind so genannte *Intrusion Detection Systeme* basierend auf Protokollen. Damit kann ein Angriff geprüft und ein Angreifer zurückverfolgt werden.

Tabelle 15: Vertraulichkeit (exemplarisch).

Anforderung (Soll)	Sicherheitsmechanismus	Zustand (Ist)
Zugriffsregelung	Authentifizierung	
Datenschutz	Verschlüsselung	
Urheberrechte	DRM	
	Digitale Wasserzeichen	
Nutzungsrechte	Authentifizierung	

Tabelle 15 beinhaltet die Sicherheitsmechanismen, die den speziellen Soll-Anforderungen für den Sicherheitsaspekt Vertraulichkeit zugeordnet sind. Zu diesen Anforderungen zählen zunächst die Zugriffsregelungen, die durch Authentifizierungen, wie sie innerhalb der Authentizität (Tabelle 14) beschrieben wurden, abgesichert werden sollten. Dies betrifft die Authentifizierung des Nutzers (Konsument), des Produzenten und der Administration, gegenüber dem Archiv.

Für die Gewährleistung des Datenschutzes kann eine Verschlüsselung der Kommunikation mit dem Archiv zum Einsatz kommen. Daten dürfen unautorisierten Benutzern nicht zugänglich sein.

Im Zusammenhang mit der Vertraulichkeit zählen darüber hinaus die Einhaltung der Urheberrechte und Nutzungsrechte zu den unabdingbaren Soll-Anforderungen an ein vertrauenswürdiges und abgesichertes digitales Langzeitarchiv. Hier ist das digitale Rechtemanagement (DRM) eine technische Maßnahme, mit welcher die Einhaltung von Nutzungsrechten erzwungen wird. Die Beachtung von Urheber- und Nutzungsrechten ist gerade bei der Bereitstellung wichtig. DRM ermöglicht die sichere Bindung von Rechtedaten an den Inhalt. Urheberrechte können mittels digitaler Wasserzeichen an ein digitales Objekt gekoppelt bzw. in das Objekt eingebunden werden.

Tabelle 16: Nachweisbarkeit (exemplarisch).

Anforderung (Soll)	Sicherheitsmechanismus	Zustand (Ist)
Originalität	s. Authentizität	
	s. Integrität	
Nachweisliche Auslieferung an autorisierte Empfänger	Protokolle	
	s. Vertraulichkeit	
	s. Verfügbarkeit	
Nachweis Archivierung	Protokolle	

Tabelle 16 beinhaltet die Sicherheitsmechanismen, die den speziellen Soll-Anforderungen für den Sicherheitsaspekt Nachweisbarkeit zugeordnet sind. Als Soll-Anforderung ist hier zunächst die Originalität aufgeführt, die je nach zu Archiv und Art der zu archivierenden Information erhalten bleiben muss. In diesem Zusammenhang ist auf die Integrität und Authentizität verwiesen, da sich die Nachweisbarkeit der Originalität von Archivobjekten auf Anforderungen und Sicherheitsmechanismen beider Sicherheitsaspekte bezieht. Eine weitere Soll-Anforderung ist die nachweisliche Auslieferung von Archivobjekten an autorisierte Empfänger. Dies kann über Protokolle erzielt werden, hier spielen aber ebenso die Sicherheitsaspekte Vertraulichkeit und Verfügbarkeit eine Rolle. Protokolle dienen gleichzeitig auch dem Nachweis der Archivierung, dass ein Objekt in das richtige Archiv aufgenommen wurde und an welcher Stelle.

Tabelle 17: Sonstige (exemplarisch).

Anforderung (Soll)		Sicherheitsmechanismus	Zustand (Ist)
Erweiterbarkeit	Speicherkapazität	Erweiterbare Systeme	
	Integration weiterer Bestände	Schnittstellen	
Interoperabilität		Standards	
Plattformunabhängigkeit		Modularisierung	
Flexibilität		Trennung Archivobjekt und Darstellung	
Skalierbarkeit		Modularisierung	
Normen/ Standards		Einheitliche Standards	
		Plattformunabhängig	
Gesetzliche Vorschriften		Datenbankenverknüpfungen	
Objekt-Formate		Einheitlich	
		Plattformunabhängig	
		Geringe Anzahl verschiedener Formate	
		Erweiterbar	
Metadaten-Formate		Textbasierte Metadaten-formate	
		Markup-Metadatenformate	
		Erweiterbar	
		Konsistent	
		Änderungen anpassend	
		Verweisend	
		Persistent	

Dokumentation	Metadaten
Transparenz	Dokumentation
Lesbarkeit für Maschine	Formate
	Metadaten
	Standards
	Migration
	Emulation
Lesbarkeit für Menschen	Darstellungsanwendung
Wirtschaftlichkeit	Kompromiss zwischen allen Anforderungen

In Tabelle 17 sind allgemeine, exemplarische Soll-Anforderungen an eine vertrauenswürdige und abgesicherte digitale Langzeitarchivierung aufgeführt und mögliche Sicherheitsmechanismen sind zugeordnet. Die in einem digitalen Archiv enthaltene Information in Form von Archivobjekten muss für die Maschine und vor allem letzten Endes für den Menschen lesbar bzw. wahrnehmbar sein. Handelt es sich um ein Langzeitarchiv, muss dies über einen langen Zeitraum sichergestellt sein. Dabei soll das Archiv einfach und transparent aufgebaut sowie effektiv handhabbar sein, gleichzeitig soll es aber auch wirtschaftlich sein. Ein digitales Langzeitarchiv muss zudem abgesichert sein gegenüber Manipulationen und Angriffen. Die Bestandserhaltung gestaltet sich hier umfangreicher und anders als in herkömmlichen Archiven. Anders dahingehen, dass ein Archivobjekt nicht einmal aufgenommen und an einem bestimmten Ort für eine lange Zeit abgelegt ist, sondern dass die digitale Langzeitarchivierung als ein dynamischer Prozess zu sehen ist, der die ständigen Veränderungen und technischen Weiterentwicklungen berücksichtigen muss. Daher gelten als Soll-Anforderungen die Erweiterbarkeit vor allem der Speicherkapazität, die Anpassbarkeit mittels geeigneter Schnittstellen, um weitere Bestände einzubinden, die Interoperabilität und Plattformunabhängigkeit, die Flexibilität und Skalierbarkeit. Die modulare Gestaltung eines Archivs ist unumgänglich.

Ein digitales Langzeitarchiv muss die gesetzlichen Vorschriften berücksichtigen und Mechanismen bereitstellen, diese einzubinden. Solche Sicherheitsmechanismen können in Form von Datenbankverknüpfungen vorhanden sein, die sowohl Urheberrechte als auch Nutzungsrechte mit einem digitalen Archivobjekt verbinden.

In der Realisierung digitaler Langzeitarchive ist man ständig damit konfrontiert, einen Kompromiss zwischen den aufgestellten Anforderungen zu finden, denn diese sind oft konkurrierend zueinander und können nicht immer gleichzeitig stark erfüllt werden. Hier müssen dann Prioritäten gesetzt werden. Um die Lesbarkeit für den Menschen sicherzustellen, müssen aufgrund der Schnelllebigkeit der Technik oft Mechanismen angewandt werden, welche die Information in ihrer ursprünglichen Form nicht erhalten können. Ebenso verhält es sich bei der Rettung bzw. Restauration von Information, die hinsichtlich der verwendeten Datenträger auf der physischen Ebene, hinsichtlich der verwendeten Datenmodelle, -schemata und –formate auf der logischen Ebene und hinsichtlich des für die Kodierung verwendeten Hintergrundwissens auf der konzeptuellen bzw. semantischen Ebene vor dem Zerfall stehen bzw. wo deren Alterungserscheinungen in diesen Dimensionen schon fortgeschritten sind. Dann geht es um die Erhaltung des Inhalts, die eigentliche Substanz, dem Teil der Information, der einen Mehrwert für die Gesellschaft hat und nicht verloren gehen darf.

Metadaten sind in der digitalen Archivierung der Schlüssel für eine erfolgreiche Umsetzung vertrauenswürdiger Archive. Metadaten sind in einem digitalen Archiv notwendig für die Auswahl, die Katalogisierung, die Klassifizierung, das Wiederfinden, das Zugreifen und das Lesbarmachen von Archivobjekten. Für die digitale Langzeitarchivierung sind darüber hinaus haltbare Metadaten unabdingbar, denn sonst ist man der Gefahr ausgesetzt, dass Archivobjekte nicht mehr gefunden, lokalisiert bzw. identifiziert und dargestellt werden können. Solche Metadaten müssen zunächst generell das Format der Speicherung beinhalten und/ oder Information bereitstellen, welche die Transformation beschreiben, die notwendig ist, um das Objekt in ein zum gegenwärtigen Zeitpunkt maschinen- und menschenlesbares Format umzuwandeln.

Weiterhin bleibt der Ort der Speicherung der Metadaten offen. Metadaten sollten zusammen mit dem digitalen Objekt gespeichert sein und zusätzlich an mindestens einem, vom Archivobjekt unabhängigen Ort mit zusätzlichem Backup. Bezüglich der Metadaten gibt es verschiedene Ansätze, unter anderem:

- URI – Uniform Resource Identifier
- URN – Uniform Resource Name: Teil der URI, der sich auf die Namensgebung bezieht und zum einen Namensgebungsverfahren und zum anderen robuste Resolutionsverfahren – Name interpretieren und dazugehöriges Objekts lokalisieren – einschließt, muss global eindeutig und skalierbar sein
- URL – Uniform Resource Locator: Identifiziert Objekt über Zugriffscharakteristik, in der die Information über den Speicherort mitcodiert ist.
- DNS – Domain Name System: Bildet Domainnamen einer URL auf IP-Adresse eines Servers ab und überträgt dadurch eine Anfrage auf einen Dienst, der den Zugriff auf das Objekt ermöglicht.
- PI – Persistenter Identifikator
- METS – Metadatenstandard
- PREMIS
- LMER
- DC – Dublin Core: Standard ISO 15836
- BMF – Broadcast Metadata Exchange Format
- FESAD –Datenbankmodell ARD-Fernseharchive
- MPEG-7
- MPEG-21

Verwendung von *Standards und Normen* ist für eine Langzeitarchivierung digitaler Information von zentraler Bedeutung, denn nur so können Anforderungen wie Interoperabilität, Skalierbarkeit oder Plattformunabhängigkeit erreicht werden. Zu solchen Standards zählen unter anderem:

- ISO 14721 OAIS
- ISO 15489 Records Management

Ebenso sollten *Formate* benutzt werden, die plattformunabhängig interpretierbar sind und die Interoperabilität eines digitalen Langzeitarchivs unterstützen. Zu solchen Formaten zählen:

- Text: PDF, RDF
- Bild: JPEG, TIFF
- Audio: WAV, MP3
- 3D-Grafik: VRML
- Video: MPEG1-4
- Metadaten: Markup Sprachen wie XML

Herangehensweise der Sicherheitsanalyse

Sicherheitsmechanismen sind in bestimmten Zusammenhängen konkurrierend und wirken sich teilweise dahingehend aus, dass sie die Anforderungen aufheben. Zum Beispiel Kryptographie mit Verschlüsselungstechnologien dient der Sicherheit der Integrität, kann aber in einem Langzeitarchiv dazu führen, dass digitale Archivobjekte nicht mehr verfügbar sind, nämlich dann, wenn der Schlüssel nicht mehr verfügbar ist und folglich auf das Objekt nicht zugegriffen werden kann. Solche Aufgaben, wie die Sicherung der Schlüssel über eine lange Zeit gehören beispielsweise zu den Aktivitäten der TCG.

Es kann vorkommen, dass Sicherheitsmechanismen zum Schutz eines Sicherheitsaspektes einen oder mehrere andere Sicherheitsaspekte verletzen. Mechanismen zur Bestandserhaltung wie Migration mit

dem Ziel, dass ein digitales Objekt auch in Zukunft verfügbar ist, bewirken beispielsweise gleichzeitig, dass die Originalität nicht mehr gewährleistet ist und die Integrität mindestens auf physischer und logischer Ebene ebenfalls verletzt ist.

Neben den technischen Aspekten, haben darüber hinaus organisatorische Mängel ebenso negativen Einfluss auf die Verstrauenswürdigkeit und Sicherheit in digitalen Langzeitarchiven, wie beispielsweise die Nichtbeachtung rechtlicher Aspekte.

In den Tabellen ist eine allgemeine Auswahl von Sicherheitsmechanismen zusammengestellt, die für alle Archivsysteme zur Langzeitarchivierung anwendbar sind. Ausgangspunkt waren die Aspekte Verfügbarkeit, Integrität, Authentizität, Vertraulichkeit und Nachweisbarkeit, die in IT-Systemen, wie es digitale Langzeitarchivierungssysteme sind, die Sicherheit festlegen. Nur mit einer angemessenen Berücksichtigung dieser Sicherheitsaspekte lassen sich digitale Langzeitarchivierungssysteme vertrauenswürdig und abgesichert gestalten.

In Abhängigkeit von Design, Konfiguration und Implementierung jedes einzelnen Langzeitarchivsystems sind die aufgelisteten Anforderungen als auch der umgesetzte und vorhandene Ist-Zustand verschieden. Anforderungen für eine vertrauenswürdige und abgesicherte Langzeitarchivierung sollten für jedes Archiv in einer *Security Policy* festgehalten sein.

Diese Expertise bezieht sich auf die Sicherheit

a) der *Systemkomponenten*, ausgehend von der Erhebung der Systemarchitektur, sowie
b) des *digitalen Archivobjektes*, ausgehend von der Erhebung der Informationsflüsse.

Die Sicherheitsanalyse geschieht demnach zum einen systemorientiert und zum anderen objektorientiert und zeigt dort den Handlungsbedarf auf. Eine szenarienabhängige und systemspezifische Analyse würde eine Netzwerksicherheitsanalyse mit expliziten und detaillierten Security-Scans erfordern. Dies war in dieser Studie nicht vorgesehen. Ausgehend von der spezifischen Erhebung der Systemarchitektur und der Informationsflüsse der zwei verschiedenartigen Beispielszenarien von Hochschul-Medienzentren und Rundfunkanstalten wurde für eine Abstraktion der anwendbaren Sicherheitsmechanismen über diese Anwendungsszenarien hinweg vollzogen. Dadurch lassen sich allgemeingültige Aussagen treffen über einen generellen Handlungs- und Standardisierungsbedarf: Denn Ziel dieser Studie war es, den bestehenden nestor-Kriterienkatalog, der allgemeingültig ist für alle Arten von digitalen Langzeitarchiven, im Punkt Sicherheit zu erweitern und Vorschläge für zukünftige Weiterentwicklungen zu präsentieren.

5.2 Reflektion in Bezug auf die organisatorischen und technischen Rahmenbedingungen – Ist-Zustand und Soll-Anforderungen innerhalb der betrachteten Szenarien

In diesem Abschnitt wird der in der Expertise festgestellte Ist-Zustand jedes der zwei Beispielszenarien öffentlich-rechtliche Rundfunkanstalten und Hochschul-Medienzentren aufgeführt. Die Darstellung folgt den zuvor entwickelten Tabellenschemata mit der Zuordnung der allgemeinen Sicherheitstechnologien zu den Anforderungen. Dadurch können die Anforderungen und Technologien in Bezug auf die organisatorischen und technischen Rahmenbedingungen für das jeweilige Szenario reflektiert werden. Im darauf folgenden Abschnitt werden die Feststellungen beispielhaft an praktischen Beispielen validiert.

Tabelle 18: Legende für Darstellung des Ist-Zustands.

✓	Soll-Anforderung wird erfüllt, Sicherheitsmechanismus wird angewandt
-	Soll-Anforderung wird nicht erfüllt, Sicherheitsmechanismus wird nicht angewandt
	Offen, nicht eindeutig bekannt

5.2.1 Öffentlich-rechtliche Rundfunkanstalten

Diese Betrachtung bezieht sich ausschließlich auf die digitalen Rundfunkarchive.

Tabelle 19: Verfügbarkeit (exemplarisch).

Anforderung (Soll)		Sicherheitsmechanismus	Zustand (Ist)
Digitales Archivobjekt	Wiederauffindbarkeit	Identifikatoren	✓
		Index	✓
		Suchfunktion	✓
	Zugriff Versionen	Versionsverwaltung	-
		Historie	-
	Vorhandensein	Antivirenprogramm	-
	Interpretierbarkeit	Darstellungsanwendung	✓
Systemressourcen (Komponenten)	Hardware-Netzwerkaufbau	Grid	-
		Verteilte Speicherlösungen	✓
	Software - Netzwerk-verbindungen	Zugänge	✓
		Clients	✓
		Firewall	✓
		Virtual Private Network (VPN)	✓
Darstellungsanwendungen		Metadaten	✓
		Emulationen	-

Tabelle 20: Integrität (exemplarisch).

Anforderung (Soll)			Sicherheitsmechanismus	Zustand (Ist)
Digitales Archivobjekt	Physische Ebene		Migration	-
			Umkopieren	✓
			Spiegelung	✓
			Backup	✓
			Verschlüsselung - Kryptographie	-
			Hashfunktionen	✓
			Antivirenprogramm	-
	Logische Ebene		Migration	-
			Transformation	-
			Digitales Wasserzeichen	-
			Verschlüsselung – Kryptographie	-
			Hashfunktionen	-
			Antivirenprogramm	-
	Konzeptuelle Ebene	Unveränderter Inhalt	Statistische Auswertungen	
			Objektive Auswertungen	
			Subjektive Auswertungen - Wahrnehmungstests	

			Antivirenprogramm	-
Systemressourcen (Komponenten und Netzwerk)	Hardware	Unversehrtheit	Zugangsbeschränkung	✓
			„Käfige“	
			Trusted Computing (TCP), NTCB	
	Software	Unversehrtheit	Intrusion Detection System (IDS)	
			Hashfunktionen – Fingerabdrücke c) Online (*Code*) d) Offline (*Dump*)	
			Forensik	
			Antivirenprogramm	
Darstellungs-anwendungen	Unversehrtheit		Hashfunktionen – Fingerabdrücke c) Online (*Code*) d) Offline (*Dump*)	
			Forensik	
			Antivirenprogramm	

Tabelle 21: Authentizität (exemplarisch).

Anforderung (Soll)	**Sicherheitsmechanismus**	**Zustand (Ist)**
Vorbeugen Änderungen – Erhaltung Original	Emulation	-
	Antivirenprogramm	-
	Firewall	✓
Nachweis unverändertes Original	Einfache Signatur	-
	Fortgeschrittene Signatur	-
	Qualifizierte Signatur	-
	Hashfunktionen	-
	Digitale Medienforensik	
Nachweis Änderungen	Protokolle	
	Versionsverwaltung	-
	Digitale Wasserzeichen	-
	Digitale Medienforensik	
Nachweis Autor	Digitale Signatur	-
	Digitales Wasserzeichen	-
Nachweis Quelle	Digitale Medienforensik	
Nachweis Archivzeitpunkt	Zeitstempel	✓
	Digitale Signatur	-
	Digitale Medienforensik	
	Digitale Computerforensik	
Nachweis Archiv	Registrierung	
	Protokolle	

Authentifizierung	Administration	Kennwörter	✓
	Produzent	Schlüssel	✓
	Konsument	Biometrie	
	System-komponenten (Server)	IP, Protokolle	✓
	Angreifer	Intrusion Detection System (IDS)	
		Protokolle	✓
		Computerforensik	

Tabelle 22: Vertraulichkeit (exemplarisch).

Anforderung (Soll)	Sicherheitsmechanismus	Zustand (Ist)
Zugriffsregelung	Authentifizierung	✓
Datenschutz	Verschlüsselung	-
Urheberrechte	Digitales Rechtemanagement (DRM)	-
	Digitale Wasserzeichen	✓
Nutzungsrechte	Authentifizierung	✓

Tabelle 23: Nachweisbarkeit (exemplarisch).

Anforderung (Soll)	Sicherheitsmechanismus	Zustand (Ist)
Originalität	s. Authentizität	s.o.
	s. Integrität	s.o.
Nachweisliche Auslieferung an autorisierte Empfänger	Protokolle	-
	s. Vertraulichkeit	s.o.
	s. Verfügbarkeit	s.o.
Nachweis Archivierung	Protokolle	-

Tabelle 24: Sonstige (exemplarisch).

Anforderung (Soll)		Sicherheitsmechanismus	Zustand (Ist)
Erweiterbarkeit	Speicherkapazität	Erweiterbare Systeme	✓
	Integration weiterer Bestände	Schnittstellen	✓
Interoperabilität		Standards	✓
Plattformunabhängigkeit		Modularisierung	✓
Flexibilität		Trennung Archivobjekt und Darstellung	✓
Skalierbarkeit		Modularisierung	✓
Normen/ Standards		Einheitliche Standards	
		Plattformunabhängig	✓
Gesetzliche Vorschriften		Datenbankenverknüpfungen	✓

Objekt-Formate	Formate	Bild: JPEG Video: DV25/50, MPEG, WM9, MXF
	Einheitlich	-
	Plattformunabhängig	✓
	Geringe Anzahl verschiedener Formate	✓
	Erweiterbar	✓
Metadaten-Formate	Formate	FESAD, Eigenentwicklungen
	Textbasierte Metadatenformate	
	Markup-Metadatenformate	
	Erweiterbar	✓
	Konsistent	✓
	Änderungen anpassend	✓
	Verweisend	-
	Persistent	-
Dokumentation	Metadaten	✓
Transparenz	Dokumentation	
Lesbarkeit für Maschine	Formate	
	Metadaten	
	Standards	
	Migration	
	Emulation	
Lesbarkeit für Menschen	Darstellungsanwendung	
Wirtschaftlichkeit	Kompromiss zwischen allen Anforderungen	✓

5.2.2 Hochschul-Medienzentren

Diese Betrachtung bezieht sich ausschließlich auf die derzeit in Hochschul-Medienzentren angewandten Systeme zur digitalen Langzeitarchivierung. Der Abschnitt basiert auf den Ausführungen zur allgemeinen Charakterisierung der digitalen Langzeitarchivierungssysteme in Hochschul-Medienzentren und der Erhebung von wesentlichen Kenngrößen in Kapitel 3, insbesondere die Systemabstraktion mit Zuordnung der Anforderungen und Annahmen für die Vertrauenswürdigkeit und Sicherheit der angewandten Langzeitarchivierungssysteme in Kapitel 4.

Einige in traditionellen analogen Archiven fest etablierte Sicherheitsmechanismen sind in den digitalen Archiven – bislang noch – nicht integriert. Dies betrifft beispielsweise die Migration der Archivobjekte auf logischer Ebene oder die subjektive Auswertung mittels Wahrnehmungstests zur Prüfung der Integrität des Archivobjektes auf konzeptueller Ebene. Vor allem sind es Sicherheitsmechanismen, die entweder nicht wirtschaftlich, schwer umsetzbar, zu aufwändig, unzuverlässig oder schlicht nicht notwendig erscheinen.

Tabelle 25: Verfügbarkeit (exemplarisch).

Anforderung (Soll)	Sicherheitsmechanismus	Zustand

			(Ist)
Digitales Archivobjekt	Wiederauffindbarkeit	Identifikatoren	✓
		Index	✓
		Suchfunktion	✓
	Zugriff Versionen	Versionsverwaltung	-
		Historie	-
	Vorhandensein	Antivirenprogramm	-
	Interpretierbarkeit	Darstellungsanwendung	✓
Systemressourcen (Komponenten)	Hardware-Netzwerkaufbau	Grid	-
		Verteilte Speicherlösungen	✓
	Software - Netzwerk-verbindungen	Zugänge	✓
		Clients	✓
		Firewall	✓
		Virtual Private Network (VPN)	✓
Darstellungsanwendungen		Metadaten	✓
		Emulationen	✓(z. Teil)

Tabelle 26: Integrität (exemplarisch).

Anforderung (Soll)			**Sicherheitsmechanismus**	**Zustand (Ist)**
Digitales Archivobjekt	Physische Ebene		Migration	✓
			Umkopieren	✓
			Spiegelung	✓
			Backup	✓
			Verschlüsselung - Kryptographie	-
			Hashfunktionen	-
			Antivirenprogramm	-
	Logische Ebene		Migration	✓
			Transformation	✓
			Digitales Wasserzeichen	-
			Verschlüsselung – Kryptographie	-
			Hashfunktionen	-
			Antivirenprogramm	-
	Konzeptuelle Ebene	Unveränderter Inhalt	Statistische Auswertungen	-
			Objektive Auswertungen	-
			Subjektive Auswertungen - Wahrnehmungstests	-
			Antivirenprogramm	-
Systemressourcen (Komponenten und	Hardware	Unversehrtheit	Zugangsbeschränkung	✓
			„Käfige“	-

Netzwerk)			Trusted Computing (TCP), NTCB	
	Software	Unversehrtheit	Intrusion Detection System (IDS)	-
			Hashfunktionen – Fingerabdrücke e) Online (*Code*) f) Offline (*Dump*)	-
			Forensik	-
			Antivirenprogramm	-
Darstellungs-anwendungen	Unversehrtheit		Hashfunktionen – Fingerabdrücke e) Online (*Code*) f) Offline (*Dump*)	-
			Forensik	-
			Antivirenprogramm	-

Tabelle 27: Authentizität (exemplarisch).

Anforderung (Soll)		**Sicherheitsmechanismus**	**Zustand (Ist)**
Vorbeugen Änderungen – Erhaltung Original		Emulation	✓
		Antivirenprogramm	-
		Firewall	✓
Nachweis unverändertes Original		Einfache Signatur	-
		Fortgeschrittene Signatur	-
		Qualifizierte Signatur	-
		Hashfunktionen	-
		Digitale Medienforensik	-
Nachweis Änderungen		Protokolle	✓
		Versionsverwaltung	
		Digitale Wasserzeichen	-
		Digitale Medienforensik	-
Nachweis Autor		Digitale Signatur	-
		Digitales Wasserzeichen	-
Nachweis Quelle		Digitale Medienforensik	-
Nachweis Archivzeitpunkt		Zeitstempel	✓
		Digitale Signatur	-
		Digitale Medienforensik	-
		Digitale Computerforensik	-
Nachweis Archiv		Registrierung	
		Protokolle	
Authentifizierung	Administration	Kennwörter	✓
	Produzent	Schlüssel	✓
	Konsument	Biometrie	

	System-komponenten (Server)	IP, Protokolle	✓
	Angreifer	Intrusion Detection System (IDS)	-
		Protokolle	
		Computerforensik	-

Tabelle 28: Vertraulichkeit (exemplarisch).

Anforderung (Soll)	Sicherheitsmechanismus	Zustand (Ist)
Zugriffsregelung	Authentifizierung	✓
Datenschutz	Verschlüsselung	-
Urheberrechte	Digitales Rechtemanagement (DRM)	
	Digitale Wasserzeichen	
Nutzungsrechte	Authentifizierung	✓

Tabelle 29: Nachweisbarkeit (exemplarisch).

Anforderung (Soll)	Sicherheitsmechanismus	Zustand (Ist)
Originalität	s. Authentizität	s.o.
	s. Integrität	s.o.
Nachweisliche Auslieferung an autorisierte Empfänger	Protokolle	
	s. Vertraulichkeit	s.o.
	s. Verfügbarkeit	s.o.
Nachweis Archivierung	Protokolle	

Tabelle 30: Sonstige (exemplarisch).

Anforderung (Soll)		Sicherheitsmechanismus	Zustand (Ist)
Erweiterbarkeit	Speicherkapazität	Erweiterbare Systeme	✓
	Integration weiterer Bestände	Schnittstellen	✓
Interoperabilität		Standards	✓
Plattformunabhängigkeit		Modularisierung	✓
Flexibilität		Trennung Archivobjekt und Darstellung	✓
Skalierbarkeit		Modularisierung	✓
Normen/ Standards		Einheitliche Standards	
		Plattformunabhängig	✓
Gesetzliche Vorschriften		Datenbankenverknüpfungen	✓
Objekt-Formate		Formate	Bild: JPEG, TIFF Text: PDF

	Einheitlich	✓
	Plattformunabhängig	
	Geringe Anzahl verschiedener Formate	✓
	Erweiterbar	✓
Metadaten-Formate	Formate	METS XML
	Textbasierte Metadaten-formate	
	Markup-Metadatenformate	
	Erweiterbar	✓
	Konsistent	✓
	Änderungen anpassend	✓
	Verweisend	-
	Persistent	-
Dokumentation	Metadaten	✓
Transparenz	Dokumentation	
Lesbarkeit für Maschine	Formate	✓
	Metadaten	✓
	Standards	
	Migration	✓
	Emulation	✓
Lesbarkeit für Menschen	Darstellungsanwendung	✓
Wirtschaftlichkeit	Kompromiss zwischen allen Anforderungen	✓

5.3 Validierung der Einsetzbarkeit an praktischen Beispielen – Ist-Zustand

In diesem Abschnitt wird der Ist-Zustand der Tabellen an praktischen Beispielen validiert und die Einsetzbarkeit der Sicherheitsmechanismen beschrieben.

5.3.1 Öffentlich-rechtliche Rundfunkanstalten

Verfügbarkeit

Kontinuität im Sendebetrieb ist für Rundfunkanstalten eine Anforderung von herausragendem Stellenwert. In der vernetzten Produktionsumgebung verlangt dies letztlich auch von dem digitalen Rundfunkarchiv sowohl eine hohe Verfügbarkeit von Systemressourcen, Darstellungsanwendungen als auch der eigentlichen multimedialen Inhalten. Entsprechend ist die technische Infrastruktur ausgelegt; *Zugänge* und *Clients* sind nahezu vollständig redundant vorhanden, *Firewalls* trennen kritische Bereiche ab und *VPN*-Lösungen binden Außenstandorte an. Havarielösungen ermöglichen einen Notbetrieb [WiSa03].

Besonders ausgeprägt ist die verteilte Speicherung der Inhalte. Durch die gesamte vernetzte Produktionsumgebung zieht sich eine Speicherhierarchie von der Online-Ebene unmittelbar in der Produktion bis hin zur Archiv-Ebene innerhalb des Rundfunkarchivs. Bemerkenswerterweise verzichten Rundfunkanstalten in den meisten Fällen auf eine Externalisierung/ Auslagerung von Datenbändern. Eine Offline-Speicherebene ist damit nicht vorhanden, allenfalls werden Überlaufbänder aus den in ihrer Kapazität ausgeschöpften Bandrobotikmagazinen genommen. Andererseits hat

eine Rundfunkanstalt aus Sicherheitsgründen ihre zwei Archivanlagen nachträglich räumlich voneinander getrennt. Verteilte-Speichersysteme im Sinne von Grid-Storage kommen massiv zum Einsatz. Von einer verteilten Speicherung mittels *Grid*-Computing ist der Rundfunk in der Anwendung jedoch noch entfernt. Eine *verteilte Speicherung* von Medieninhalten über mehrere gleichberechtigte Standorte hinweg stellt dagegen kein grundsätzliches Problem dar. Wie jedoch eine im Ausland bereits praktizierte Auslagerung der Archivobjektspeicherung an externe Dienstleister bezüglich Verfügbarkeit zu werten ist soll hier offen gelassen werden.

Die Bestände der Rundfunkarchive werden mit gutem Grund auch als Programmvermögen bezeichnet. Fernsehproduktionen sind kostspielig und ihre Wiederverwendung ist wirtschaftlich attraktiv – ganz abgesehen von der kulturellen Bedeutung vieler Archivobjekte des öffentlich-rechtlichen Rundfunks. Für die hohe Wiederverwertungsrate ist *Wiederauffindbarkeit* bedeutsam. Archivobjekte werden formal und inhaltlich erschlossen, im *Index* zunehmend detailliert erfasst und die Bestände sind mittels *Suchfunktionen* vielfältig recherchierbar, auch über Archivgrenzen hinweg.

Während die formale Erschließung strengen Richtlinien unterworfen ist und Mindestangaben abdeckt, ist die inhaltliche Erschließung subjektiv geprägt in ihrer Form, ihrem Umfang und ihrer Wortwahl. Die Ausrichtung und Zielsetzung der Rundfunkanstalt beeinflusst maßgeblich, inwieweit Sachinhalte, Bildinhalte und Rechte bei der inhaltlichen Erschließung eingearbeitet werden.

Die eindeutige Kennzeichnung von Medieninhalten ist eine Selbstverständlichkeit; weil Metadaten und Essenz nur hierüber zugeordnet werden können sind *Identifier* sogar unverzichtbar.

Darstellungsanwendungen schließen im Rundfunk grundsätzlich auch sämtliche technischen Geräte zur Bearbeitung oder Ausstrahlung, also beispielsweise Videoschnittplätze und Videoserver, mit ein. Diese sind in den Rundfunkanstalten in Vielzahl vorhanden. Gemeinsame Nenner für die Medienformate in dieser heterogenen Infrastruktur zu finden ist nicht trivial. Vor diesem Hintergrund wird eine Umstellung der Formate voraussichtlich eher selten vorgenommen. Trotzdem kennzeichnen technische *Metadaten* das Format von Archivobjekten.

Integrität

Die öffentlich-rechtlichen Rundfunkarchive sind sehr auf die Integrität ihrer Archivalien bedacht. Auf physischer Ebene kommen bewährte informationstechnische Verfahren zur Datenspeicherung zum Einsatz. Die digitalen Archive betreiben redundant ausgelegte, verteilte Speichersysteme, in denen auf physischer Ebene die digitalen Archivobjekte mittels *Spiegelung* und *Backup* mehrfach vorgehalten werden. *Hashfunktionen* prüfen die Unversehrtheit der Dateien. Per *Umkopieren* werden Archivdateien von als potenziell schadhaft erkannten Datenbändern genommen.

Ein manueller Qualitätssicherungsprozess der digital archivierten Daten innerhalb der digitalen Archive ist keine Selbstverständlichkeit. Automatisierte Prüfungsverfahren für die Qualität von Videoinformation existieren am Markt, kommen bei der Bestandssicherung im Archiv jedoch derzeit eher nicht zum Tragen. Der Aufwand in Form von Infrastrukturkosten und Latenzzeiten derartiger Analysen wird gescheut. Bei der Rückwärtsdigitalisierung – insbesondere vom Zerfall bedrohten, analogen Materials – wird die Einspielung manuell überwacht und eine Qualitätssicherung ist unabdingbar.

Ein Einsatz von Sicherheitsmechanismen auf logischer als auch konzeptueller Ebene zur Prüfung der Integrität von Archivobjekten im Bestand der digitalen Rundfunkarchive ist dieser Expertise nicht bekannt, kann letztendlich aber nicht ausgeschlossen werden.

Wegen des geringen Alters der digitalen Archive der öffentlich-rechtlichen Rundfunkanstalten war eine *Migration* der Archivbestände bislang nicht zwingend erforderlich. Mit der bei einem Formatwechsel einhergehenden Transkodierung werden Qualitätseinbußen befürchtet.

Die Dateiformate für Audio- und Videomaterial und beispielsweise auch das Containerformat MXF werden als potenzielle Virenträger eingestuft. Noch ist jedoch unklar, ob gängige Antivirenprogramme derartige Formate handhaben können. Auch ist die Frage der Schnittstellen für eine Einbindung der Virenprüfung ungeklärt.

In den Rundfunkanstalten werden neu in das Langzeitarchiv eingelieferte Materialien nicht auf Virenbefall geprüft. Ebenfalls werden Archivbestände nicht auf Viren getestet. Dies begründet sich dadurch, dass eine derartige Prüfung zu aufwändig und langwierig wäre. Einerseits würde sich der

Einlieferungsvorgang verlangsamen, andererseits würden die Speichersysteme für eine Virenprüfung zur Bestandserhaltung mit zusätzlichen Lese-/ Schreibzugriffen belastet. Um beide Nachteile zu vermeiden wurde der Vorschlag aufgebracht, neue Archivalien zunächst über eine gewisse Zeit von beispielsweise 30 Tagen in Quarantäne zwischen zu lagern, bevor sie einmalig eine Prüfung durch Antivirenprogramme durchlaufen und endgültig in das Langzeitarchiv überführt werden. Eine Umsetzung derartiger Regelungen erforderte in erster Linie organisatorische Maßnahmen.

Mit deren steigender Leistungsfähigkeit bauen Rundfunkanstalten ihre technische Infrastruktur zunehmend aus angepasster Standard-Informationstechnik anstatt spezieller teuerer Gerätetechnik auf. Dadurch werden sie grundsätzlich anfällig für Schädlinge und Angriffsmuster aus der Informationstechnik. Weil Arbeitsplatzrechner oftmals sowohl Zugriff auf das Produktionsnetz als auch das weltweite Datennetz Internet haben, stellen sie potenzielle Einfalltore für Angriffe dar. Produktionsnetze in den Rundfunkanstalten sind daher in der Regel von der Vernetzung der Redaktionen separiert und die Übergänge über Firewalls geschützt. Generell kommen etablierte Verfahren der Informationstechnik zum Schutz der technischen Infrastruktur zum Tragen.

Authentizität

Bestimmendes Mittel für eine Sicherstellung der Authentizität in den öffentlich-rechtlichen Rundfunkanstalten ist der Schutz der Infrastruktur und des gesicherten, authentifizierten Zugangs zu dieser. Umfassende organisatorische und technische Maßnahmen schützen die Infrastruktur vor unberechtigtem Zugriff.

Die vernetzte Produktionsumgebung unterscheidet individuelle Nutzer und/ oder ihre Rollen. Rollenspezifische Berechtigungen tragen zur Einordnung als Administrator, Produzent oder Konsument gegenüber dem Archiv bei. Dass die technische Infrastruktur der Rundfunkarchive auch durch mechanische Maßnahmen wie beispielsweise sperrbare Türen geschützt ist nimmt diese Expertise als selbstverständlich an, ohne dies ausdrücklich erhoben zu haben. Ob zur *Authentifizierung* berechtigter Nutzer *Kennwörter*, mechanische *Schlüssel* oder *biometrische Merkmale* herangezogen werden sei an dieser Stelle nachrangig.

Weniger ausgeprägt ist hingegen der nachträgliche, belastbare Nachweis von Authentizität, insbesondere unmittelbar bezogen auf das individuelle Archivobjekt.

Ein Einsatz von *Digitalen Signaturen*, von *Digitaler Medienforensik* oder von *Digitaler Computerforensik* wurde in der Erhebung nicht festgestellt, kann letztlich aber auch nicht ausgeschlossen werden. *Digitale Wasserzeichen* kommen zwar durchaus zum Einsatz, jedoch nicht zum Nachweis der Authentizität eines Archivobjektes sondern vielmehr zum Zweck der Markierung urheberrechtlich geschützten Materials.

Digitale Rundfunkarchive erfassen den Zeitpunkt der Einlieferung von Archivalien.

Vertraulichkeit

Administratoren, Produzenten als auch Konsumenten im Sinne des Referenzmodells authentifizieren sich gegenüber der vernetzten Produktionsumgebung und auch gegenüber dem darin eingebetteten digitalen Rundfunkarchiv. Das Archivmanagement regelt Zugriffsrechte auf Archivobjekte anhand der Gruppenzugehörigkeit des Benutzers.

Die öffentlich-rechtlichen Rundfunkanstalten haben sich stets gegen eine *Verschlüsselung* bei der Sendeausstrahlung zum Konsumenten ausgesprochen, sie verfolgen das Modell des freien Empfangs. Konsequenterweise wird bei der Verbreitung kein *Digitales Rechtemanagement (DRM)* eingesetzt.

Die Verwaltung von Urheber- und Nutzungsrechten zu den Beständen in einer Sendeanstalt ist derzeit Aufgabe eigenständiger, regelmäßig Datenbank-basierter Systeme der Abteilung Honorare-&-Lizenzen (HoLi). Die Archive setzen die sich aus den Rechtedaten ergebenden Einschränkungen und Vorschriften vollständig, jedoch nicht in jedem Fall automatisiert um. Die Archivobjekte selbst sind nicht an ein DRM gebunden. Solch eine technische Zwangsmaßnahme wäre auch problematisch, denn die Produktionsgeräte der Rundfunkanstalten sind grundsätzlich nicht für einen Umgang mit DRM

ausgelegt. Allenfalls, dies betrifft Vorschaumaterial, werden Archivobjekte mit *Digitalen Wasserzeichen* (*Watermarking*) gekennzeichnet.

Nachweisbarkeit

Die Rundfunkanstalten haben grundsätzlich ein Interesse an einer Aufzeichnung sowohl der Bearbeitungsschritte als auch der Abrufe von Medienobjekten. Primärer Zweck derartiger *Protokolle* wäre eine Optimierung der Produktionsprozesse und eine Qualitätsverbesserung, erlauben sie doch eine Einschätzung der Arbeitsaufwände respektive des Wertes der Medienobjekte. Einer Protokollierung der Kommunikation sind jedoch Grenzen gesetzt. Sobald die Aufzeichnung einen Bezug auf individuelle Mitarbeiter herstellte, wären Persönlichkeitsrechte betroffen und Datenschutzrichtlinien greifen.

Zu einem *Nachweis der Auslieferung von Archivobjekten* respektive deren *Archivierung* könnten derartige Protokolle der Kommunikation des Archivs grundsätzlich herangezogen werden.

Sonstige

Die Rundfunkanstalten setzen zunehmend angepasste *Standard*-Informationstechnik anstatt spezieller teuerer Gerätetechnik ein. Damit steht für grundlegende Komponenten austauschbare Hard- und Software in großer Auswahl bereit. Ganz besonders gültig ist dies für die Netzwerktechnik, Anwendungsserver oder Arbeitsplatzrechner. Ebenfalls zutreffend, aber in geringerem Umfang, ist dies für Datenpumpen oder Videoserver. Auf Teilsystemebene ist weniger *Flexibilität* gegeben, weil diese an die Besonderheiten der Rundfunkanstalt, zu nennen sind beispielsweise Arbeitsabläufe und Metadatenschemata, umfangreich angepasst wurden.

Bei den Herstellern von Techniklösungen für den Rundfunk ist eine wachsende Bereitschaft an vereinheitlichten Lösungen zur *Interoperabilität* zu beobachten. Eine modulare, offene Systemarchitektur mit definierten Schnittstellen wurde als eine unabdingbare Forderung für die auf Informationstechnik gestützte Produktionsumgebung des Rundfunks erkannt. Die Hersteller bieten aber in allererster Linie einzelne Großsysteme an. In der Praxis entschließen sich öffentlich-rechtliche Rundfunkanstalten bei der Planung ihrer technischen Infrastruktur sehr unterschiedlich. Die Ansätze reichen vom Verbund integrierter Teilsysteme verschiedener Hersteller bis hin zu Gesamtsystemen aus einer Hand. Zur Implementierung von Verbundsystemen notwendig sind allerdings Anpassungen der Teilsysteme sowie eine Neuentwicklung zahlreicher Schnittstellenformate und Vermittlungsmechanismen zwischen den Teilsystemen. Die digitalen Archive sind modular aufgebaut. So beherbergt ein Anwendungsserver das Archivmanagement, während eine Bandstation und Plattenspeichersysteme für die Speicherung der Archivalien verantwortlich zeichnen. Diese Modularisierung innerhalb des Archivs folgt nicht zwingend dem Referenzmodell, lässt sich aber in weiten Teilen darauf abbilden. Die *Trennung von Archivobjekt und Darstellung* ist eindeutig gegeben.

Hohe Verfügbarkeit und Realzeitfähigkeit sind die wesentlichen Merkmale der IT-Lösungen für den Rundfunk. *Erweiterungen der Hard- und Softwaresysteme* müssen im laufenden Betrieb erfolgen können.

Gesetzliche Vorschriften, welche eine Verknüpfung von Datenbanken einforderten, sind der Expertise für den Bereich der öffentlich-rechtlichen Rundfunkanstalten nicht bekannt. Für die Feststellung von Urheber- und Nutzungsrechten wird für Archivobjekte technisch ein Bezug auf die Rechtedaten der Abteilung Honorare-&-Lizenzen hergestellt.

In der Regel verwenden keine zwei Rundfunkanstalten das gleiche Metadatenmodell. Eine Ausnahme sind die in der ARD organisierten Anstalten, von denen die überwiegende Zahl im FESAD-Konsortium organisiert ist und das von FESADneu angebotene Datenmodell verwenden. Über zahlreiche Eigenentwicklungen der Rundfunkanstalten und proprietäre Herstellerformate hinaus lassen sich einige wenige Metadatenschemata ausmachen, beispielsweise BMF.

Die Metadatenschemata innerhalb der einzelnen Rundfunkarchive sind in sich jedoch *konsistent*. Moderne Content-Managementsysteme erlauben eine große Gestaltungsfreiheit und legen damit die Grundlage sowohl für eine *Erweiterung* des Metadatenschemas im Archiv als auch dessen *Anpassung an Änderungen*.

Die Erhebung hat gezeigt, dass die Rundfunkanstalten die Wahl der von ihnen verwendeten Objektformate in Produktion und Archiv sorgfältig getroffen haben. Ihre *Anzahl* ist gering gehalten. *Einheitlich* sind die Objektformate letztlich aber nicht. Die begonnene Diversifizierung der vom öffentlich-rechtlichen Rundfunk bedienten Verbreitungswege lässt ebenfalls einen Anstieg der zu berücksichtigenden Formate erwarten, wenn allerdings auch nur bedingt innerhalb der Archive selbst. Bemerkbar ist dies bereits bei Vorschaumaterial. Weil die Rundfunkanstalten bemüht sind, die Medienobjekte in ihrer heterogenen technischen Infrastruktur, dies schließt unter anderem Videoserver, Videoschnittplätze und Sendeautomationen mit ein, soweit als möglich ohne Formatwechsel zu verarbeiten, sind die Objektformate in aller Regel *Plattformunabhängig* und weitestgehend standardisiert. Zu nennen sind hier beispielsweise JPEG, MPEG und das Containerformat MXF. Die Leistungsfähigkeit der Archive schränkt das Spektrum von ihnen verwaltbarer Essenzformate grundsätzlich nicht ein.

Die öffentlich-rechtlichen Rundfunkanstalten berücksichtigen bei der Planung und Durchführung von Archivierungsmaßnahmen – unter anderem, aber nicht ausschließlich – Anforderungen des Versorgungsauftrags, der Selbstverpflichtung zur Bewahrung ihres Programmvermögens, des wirtschaftlichen Wertes von Archivalien, des eigenen Produktions- und Sendebetriebes sowie des wirtschaftlichen Handelns.

5.3.2 Hochschul-Medienzentren

Verfügbarkeit

Kontinuität in der Umgebung des digitalen Langzeitarchivs in Hochschul-Medienzentren hat für die Sicherung der Verfügbarkeit einen sehr hohen Stellenwert. In einer vernetzten Umgebung betrifft die Verfügbarkeit nicht nur die digitalen Archivobjekte mit ihren multimedialen Inhalten, sondern gleichzeitig auch Systemressourcen und Darstellungsanwendungen. Dementsprechend ist die technische Infrastruktur ausgerichtet, indem *Zugänge* und *Clients* nahezu vollständig redundant vorhanden sind, *Firewalls* kritische Bereiche abtrennen und *VPN*-Lösungen Außenstandorte anbinden. Havarielösungen, sofern vorhanden, ermöglichen einen Notbetrieb.

Zum größten Teil wird die verteilte Speicherung der Inhalte angewandt bzw. angestrebt. So werden teilweise eine bzw. mehrere Kopien des Bestands an verschiedenen geografischen Orten gespeichert. Solch eine räumliche Verteilung ist jedoch nicht immer vorhanden. In der Archivumgebung ist eine Speicherhierarchie vorhanden, was bedeutet, dass ein Objekt auf mehreren Ebenen gespeichert wird bevor es in die eigentliche Ebene des Archivspeichersgelangt. In den meisten Fällen wird ein Archivobjekt mit hoher Zugriffsrate zunächst auf einem so genannten Dokumentenserver (*Cache*) gespeichert und bei geringer Zugriffsrate in die Ebene des Archivspeichers verschoben. Im Archivspeicher selbst kommen verteilte Speichersysteme im Sinne von *Grid-Storage* oft zum Einsatz. Eine Auslagerung (Externalisierung) von Datenbändern ist nicht vorhanden.

Im Zusammenhang mit der *Wiederauffindbarkeit* werden Archivobjekte formal und inhaltlich erschlossen. Sie werden im Index bzw. in entsprechenden Datenbanken erfasst. Die Bestände sind mittels *Suchfunktionen* vielfältig recherchierbar, auch über Archivgrenzen hinweg. Es werden standardisierte Identifikatoren, wie Persistent Identifier (PI), URN, URI oder URL angewandt, um ein Objekt im Archivspeicher zu lokalisieren. Die Erschließung der Inhalte in Hochschul-Medienzentren ist teilweise nicht einheitlich und Metadaten sind nicht vollständig bzw. variieren in ihren Angaben. Die Hochschul-Medienzentren sind hier bemüht, dies zu vereinheitlichen.

Darstellungsanwendungen werden vorrangig mittels technischer *Metadaten* mit dem Archivobjekt verbunden. Technische *Metadaten* kennzeichnen ebenfalls das Format von Archivobjekten. Je nach Archiv werden die Darstellungsanwendungen selbst zusätzlich archiviert und können bei Bedarf emuliert werden.

Langfristiges Vorhandensein sollte, wie in den Soll-Anforderungen festgehalten, mittels Antivirenprogrammen abgesichert sein. Dies hat sich aufgrund des erheblichen Mehraufwandes in den digitalen Langzeitarchiven der Hochschul-Medienzentren derzeit jedoch noch nicht durchgesetzt.

Integrität

Die Sicherung der Integrität wird in Hochschul-Medienzentren derzeit mittels Migration, Kopieren, Transformation, Backupstrategien sowie Zugangsbeschränkungen gelöst. Mechanismen der IT-Sicherheit wie Hashfunktionen, Antivirenprogramme, Verschlüsselungstechnologien oder digitale Wasserzeichen kommen hier bisher selten zum Einsatz. Dies sollte im Sinne der vertrauenswürdigen und abgesicherten Langzeitarchivierung ausgebaut werden.

Mechanismen wie *Hashfunktionen* oder *digitale Wasserzeichen* zur Überprüfung der Unversehrtheit der Ressourcen, Darstellungsanwendungen und nicht zuletzt des Archivobjektes kommen nicht zur Anwendung, was das Erkennen von potenziell schadhaften oder manipulierten Datenbeständen erschwert, wenn nicht sogar ausschließt. Automatisierte Prüfverfahren existieren am Markt, werden bei der Bestandssicherung jedoch aufgrund des Aufwands in Form von Infrastrukturkosten und Latenzzeiten derartiger Analysen derzeit eher nicht eingesetzt. Ein Einsatz von Sicherheitsmechanismen über Migrationen hinaus auf logischer als auch konzeptueller Ebene zur Prüfung der Integrität von Archivobjekten im Bestand der Hochschul-Medienzentren ist dieser Expertise nicht bekannt, kann letztendlich aber nicht ausgeschlossen werden.

Wegen des geringen Alters der digitalen Archive der Hochschul-Medienzentren war eine *Migration* der Archivbestände bislang nicht zwingend erforderlich. Mit der bei einem Formatwechsel einhergehenden Transkodierung werden Qualitätseinbußen befürchtet. Bisher ist keine Sicherheits-Lösung integriert, die solche Qualitätseinbußen überprüfen könnte.

Die Dateiformate für Audio- und Videomaterial werden im Allgemeinen als potenzielle Virenträger eingestuft, da es in der Vergangenheit vermehrt zu Vorkommnissen dieser Art gekommen ist. Unklar ist dabei, inwieweit Antivirenprogramme derartige Formate handhaben können und einen Virenbefall erkennen. Darüber hinaus ist die Frage der Schnittstellen für eine Einbindung der Virenprüfung ungeklärt. In den Hochschul-Medienzentren werden weder neu in das Langzeitarchiv eingelieferte Materialien auf Virenbefall geprüft noch werden Archivbestände auf Viren getestet. Derartige Prüfungen wären aufwändig und langwierig, der Einlieferungsvorgang würde sich verlangsamen und Speichersysteme würden für eine Virenprüfung zur Bestandserhaltung mit zusätzlichen Lese-/Schreibzugriffen belastet. Aufgrund der Dringlichkeit der Virenproblematik müssen hier jedoch in naher Zukunft Lösungswege gefunden werden. Mit zunehmender externer Interaktion zu beispielsweise Konsumenten oder Produzenten als auch mit der archivübergreifenden Kommunikation sind digitale Langzeitarchive zunehmend anfällig für Schadprogramme. Zudem ist auf den Arbeitsplatzrechnern oftmals gleichzeitig der Zugriff sowohl auf das interne Archivnetz als auch das weltweite Datennetz Internet erlaubt. Somit stellen sie potenzielle Einfalltore für Angriffe dar.

Authentizität

Im Sinne der vertrauenswürdigen und abgesicherten Langzeitarchivierung ist die Authentizität ähnlich wie die Integrität nur bedingt gesichert. Grundsätzlich betrifft die Sicherung der Authentizität in Hochschul-Medienzentren den Schutz der Infrastruktur mittels des gesicherten, authentifizierten Zugangs. So werden umfassende organisatorische und technische Maßnahmen erforderlich, um die Infrastruktur vor unberechtigten Zugriffen zu schützen.

Zugangsberechtigungen sind entsprechend der Nutzer und ihren Rollen als Administrator, Produzent oder Konsument gegenüber dem Archiv festgelegt. So werden zur *Authentifizierung* berechtigter Nutzer *Kennwörter* und mechanische *Schlüssel* herangezogen. Inwieweit die Verwendung oder *biometrischer Merkmale* zur Authentifizierung herangezogen werden kann und sollte, sei an dieser Stelle offen gelassen und ist je nach Archivinhalt zu beurteilen.

In Archiven der Hochschul-Medienzentren ist der nachträgliche Nachweis von Authentizität, insbesondere die des Archivobjekts, weniger vorzufinden. Ein Einsatz von *Digitalen Signaturen*, von *Digitaler Medienforensik* oder von *Digitaler Computerforensik* wurde in der Erhebung nicht festgestellt, kann letztlich aber auch nicht ausgeschlossen werden. Auch der Einsatz von *Digitalen Wasserzeichen* konnte nicht festgestellt werden.

Die Authentifizierung von Systemkomponenten im Netzwerk geschieht über ihre *IP-Adresse* und *Protokolle*.

Der *Zeitpunkt* der Einlieferung von Archivobjekten wird festgehalten, offen ist jedoch inwieweit Änderungen lückenlos dokumentiert werden.

Vertraulichkeit

Administratoren, Produzenten als auch Konsumenten im Sinne des Referenzmodells authentifizieren sich gegenüber der Archivumgebung. Das Archivmanagement regelt Zugriffsrechte auf Archivobjekte. Eine Verschlüsselung im Sinne des Datenschutzes bleibt offen. Eine Verschlüsselung der Kommunikation mit dem Archiv sollte aber für eine vertrauenswürdige und abgesicherte Langzeitarchivierung angestrebt werden. In Bezug auf die Urheberrechte ist offen, ob ein *Digitales Rechtemanagement (DRM)* oder *digitale Wasserzeichen* in Hochschul-Medienzentren Anwendung finden. Ebenfalls offen ist, ob in Bezug auf die Nutzungsrechte ein *Digitales Rechtemanagement* (DRM) als eine technische Maßnahme, mit welcher die Einhaltung von Nutzungsrechten erzwungen wird, zum Einsatz kommt. Auch die Regelung der Auslieferung von Inhalten ohne DRM an berechtigte Nutzer konnte nicht erhoben werden.

Nachweisbarkeit

Protokolle können grundsätzlich dazu herangezogen werden, um die nachweisliche Auslieferung von Archivobjekten an autorisierte Empfänger oder deren Archivierung zu sichern. Inwiefern davon derzeit Gebrauch gemacht wird, bleibt offen.

Sonstige

Generell befinden sich die digitalen Langzeitarchive der Hochschul-Medienzentren im Aufbau und sind bemüht sich an aktuelle Gegebenheiten anzupassen. Hochschul-Medienzentren setzen für Netzwerktechnik, Server oder Arbeitsplatzrechner vorrangig angepasste *Standard*-Informationstechnik ein. So sind beispielsweise die technischen Komponenten für den Archivspeicher von aktuellen Marktanbietern realisiert, während *Clients* zum *Ingest* oft selbst entwickelt sind und dabei existierende Methoden zur Erschließung einbinden. Damit stehen für grundlegende Systemkomponenten und Ressourcen austauschbare Hard- und Software bereit, was die *Erweiterbarkeit* und *Flexibilität* realisiert. Zusätzlich sind für die Integration weiterer Bestände oder neuer Archive Schnittstellen geschaffen, dies jedoch nicht standardisiert. Grundsätzlich sind vereinheitlichende Lösungen angestrebt, um eine *Interoperabilität* herzustellen. Diese betrifft die sowohl Arbeitsabläufe als auch Systemressourcen, Metadaten oder Formate. Somit soll der Forderung nach einer modularen und offenen Systemarchitektur mit definierten Schnittstellen Folge geleistet werden.

In Hochschul-Medienzentren ist man grundsätzlich bemüht die Systemarchitektur und Umgebung entsprechend des OAIS-Referenzmodells umzusetzen. Die *Trennung von Archivobjekt und Darstellung* ist eindeutig gegeben.

Gesetzliche Vorschriften, welche eine Verknüpfung von Datenbanken einforderten, sind der Expertise für den Bereich der Hochschul-Medienzentren nicht bekannt. Für die Feststellung von Urheber- und Nutzungsrechten wird für Archivobjekte technisch ein Bezug auf die organisatorisch festgelegten Verantwortlichen für Rechtedaten hergestellt.

Die verwendeten Metadatenschemata der einzelnen Hochschul-Medienzentren sind in sich *konsistent* und meist ähnlich und sind somit aufeinander abbildbar. Der Einsatz von Content-Managementsystemen ist nicht bekannt. Bezüglich der Metadaten sind die Hochschul-Medienzentren bemüht Standards wie METS oder Dublin Core einzusetzen.

In Bezug auf die verwendeten Objektformate sind Hochschul-Medienzentren darauf bedacht, keine unnötige Einschränkung in Hinblick auf die verwaltbaren Formate zu machen. Mit Blick auf die Austauschbarkeit kommen derzeit jedoch hauptsächlich die Gängigsten zum Einsatz. Eine einheitliche Verwendung von Formaten ist momentan nicht der Fall, da in jedem Archiv eigene Lösungen oft eigenständig entwickelt werden. Dies schließt gleiche Lösungen in verschiedenen Archiven aber nicht aus. Formate sind in aller Regel *Plattformunabhängig* und weitestgehend standardisiert, wie beispielsweise PDF, JPEG und MPEG.

Hochschul-Medienzentren berücksichtigen generell bei der Planung und Durchführung von Archivierungsmaßnahmen Anforderungen für eine vertrauenswürdige und abgesicherte Langzeitarchivierung, um ihrer Mission, der grundsätzlichen Erhaltung von Information von bleibendem wissenschaftlichem, künstlerischem oder gesellschaftlichem Wert gerecht zu werden. Dazu werden weiterführend verantwortliche Entwicklungen von Strategien, die den beständigen, vom Informationsmarkt verursachten Wandel bewältigen können, verfolgt. Dazu zählt unter anderem der Einsatz von standardisierten Objekt- und Metadatenbeschreibungen, aber auch die Betrachtung der Wirtschaftlichkeit. Die Betrachtung des wirtschaftlichen Wertes von Archivalien hat hier, anders als in den Rundfunkanstalten, keine vorrangige Relevanz.

Zukünftige Aktivitäten sollten vermehrt den Einsatz von Open-Source-Software im Sinne der Nachhaltigkeit evaluieren und Mechanismen der Sicherheit einbeziehen, um eine vertrauenswürdige und abgesicherte Langzeitarchivierung zu realisieren. Denn das Problem der digitalen Langzeitarchivierung ist der Aspekt der Langzeit. Insbesondere die Frage, wie mit Darstellungsanwendungen auf lange Zeit eine integere Verfügbarkeit des Inhalts sichergestellt und ein Archivobjekt für den Menschen lesbar gemacht werden kann. Die Herausforderung besteht hier in der vertrauenswürdigen Absicherung des archivübergreifenden Austausches von digitalen Objekten oder der Interaktion mit externen Produzenten sowie Konsumenten. Daraus leitet sich unmittelbar Handlungsbedarf ab.

6 Handlungsbedarf

In diesem Kapitel wird der Handlungsbedarf in Bezug auf die Gestaltung vertrauenswürdiger und abgesicherter Langzeitarchive multimedialer Inhalte aufgezeigt. Es wird dargelegt, inwieweit Sicherheitstechnologien in Langzeitarchiven integriert werden sollten bzw. dies möglich ist. Darauf aufbauend wird ein möglicher Standardisierungsbedarf abgeleitet, in denen sich das Kompetenznetzwerk Langzeitarchivierung als Ganzes engagieren und Aktivitäten initiieren muss. Dazu werden laufende Standardisierungsaktivitäten aufgeführt. Des Weiteren wird der Einsatz von potentiellen und neuartigen Technologien, wie hochgradig verteilte Speichertechnologien evaluiert, um z.B. Potentiale für den Einsatz von Grid- und anderen Virtualisierungstechnologien auf unterschiedlichen Ebenen aufzuzeigen.

6.1 Standardisierungsbedarf

Auf Grundlage der Anforderungen an die vertrauenswürdige und abgesicherte digitale Langzeitarchivierung in Kapitel 2 und der Aufstellung des Ist-Zustandes in Kapitel 5 wird in diesem Abschnitt zusammenfassend der Standardisierungsbedarf dargelegt und gleichzeitig bestehende Standards und laufende Standardisierungsbemühungen mit aktueller Relevanz aufgeführt. Standardisierungen sind in erster Linie mit den Zielen *Interoperabilität und Nachhaltigkeit* verbunden. Sie sollen den Zweck erfüllen, einen Austausch von Information zu ermöglichen und Information für eine lange Zeit unabhängig von zur Verfügung stehender Hard- und Software zugänglich zu machen. Heute entwickelte und eingesetzte Systeme sollen in Zukunft erweiterbar sein und an sich ändernde Gegebenheiten (technisch, rechtlich, kulturell, usw.) anpassbar sein. Weil Interoperabilität und Nachhaltigkeit zu den essentiellen Anforderungen der Langzeitarchivierung zählen, kommt der Standardisierung eine hohe Bedeutung zu.

Ausgehend von den Anforderungen kann ein Standardisierungsbedarf der technischen Aspekte in der digitalen Langzeitarchivierung wie folgt festgehalten werden:

1. Standardisierung von Hard- und Software sowie von (Daten-)Formaten
   - Speichersysteme
   - Netzwerkverbindungen
   - Speicherformate
   - Metadatenformate
2. Standardisierung von Architektur und Infrastruktur
   - Verteilte Systeme und Grid
3. Standardisierung von Prozessen
   - Innerhalb des Archivs – „Internal“
   - Archivübergreifend/ nach Außen – „External“
4. Standardisierung von Bestandserhaltungsstrategien
   - Migration
   - Emulation

5. Standardisierung in Bezug auf die Sicherheit
   - Sicherheitsanalyse
   - Sicherheitsmechanismen

6.1.1 Standardisierungen von Hard- und Software sowie von (Daten-)Formaten

Speichersysteme und Server

In der Regel kommen proprietäre Speichersysteme und Server von dritten Anbietern zum Einsatz.

Netzwerk

Bezüglich Netzwerke gibt es einen Standardisierungsbedarf hinsichtlich verteilter Netze und insbesondere Grid-Computing [FoKe04]. Hier sei beispielsweise die *Open Grid Service Architecture (OGSA)* erwähnt, die auf Ansätzen aus der *Open Grid Services Infrastructure (OGSI)* beruht. Auf diese Weise sollen verschiedene Komponenten als so genannte Grid-Services in einer offenen Komponentenarchitektur verbunden werden. In diesem Zusammenhang existiert die so genannte D-Grid-Initiative[30], die eine nachhaltige Grid-Computing-Infrastruktur für die erweiterte Wissenschaft (*e-Science*) in Deutschland aufbaut.

Datenbanken und Verwaltungsabläufe

In Bezug auf Datenbanken ist generell auch ein Standardisierungsbedarf vorhanden hinsichtlich der Interoperabilität und Plattformunabhängigkeit, um einen Austausch von Archivobjekten zu sichern. Die folgenden Standardisierungen im Zusammenhang mit Datenbanken und der Wiederauffindbarkeit von Archivobjekten seien hier exemplarisch benannt:

Persistente Identifikatoren sind eindeutige, standortunabhängige Identifikatoren für digitale Objekte[31]. *Format Registries* sind Datenbanken und enthalten umfassende und in standardisiert erfasster und auslesbarer Form Information zu Dateiformaten. Diese Information enthält u.a. Namen, Version, Zeichensatz und Hinweise zu Hard- und Softwareanforderungen. Hier sind Standardisierungsarbeiten erfolgt wie z.B. PRONOM[32] oder *Global Digital Format Registry (GD-FR)*[33]. *Digital Objet Identifier (DOI)*[34] dienen der Identifizierung von Inhaltsobjekten. DOI-Namen sind verbunden mit allen Entitäten, die in digitalen Netzwerken fungieren und ändern sich nicht, auch wenn sich die digitale Entität ändert.

Uniform Resource Identifier (URI), Uniform Resource Name (URN) und *Uniform Resource Locator (URL)* sind standardisiert vom W3C-Konsortium, um digitale Objekte adressieren und identifizieren zu können. Sie werden in digitalen Langzeitarchiven verbreitet verwendet. URLs sind besonders wichtig für die archivübergreifenden Referenzierung von Objekten.

Im Bereich der öffentlich-rechtlichen Rundfunkanstalten ist die etablierte Fernseharchivdatenbank (FESAD) an dieser Stelle aufzuführen, die eine Verbindung zu den auf Bild- und Tonträgern gespeicherten Essenzen herstellt.

Speicherformate

Standardisierungsbedarf hinsichtlich der Speicherformate ist dahingehend vorzuschlagen, dass sich in einem Bereich darauf geeinigt wird wenige austauschbare und möglichst gleiche Speicherformate zu verwenden, wie beispielsweise in den Rundfunkanstalten das *Material Exchange Format (MXF)*.

---

[30] http://www.d-grid.de/

[31] http://www.persistent-identifier.de/

[32] http://www.nationalarchives.gov.uk/PRONOM/default.htm

[33] http://hul.harvard.edu/gdfr/

[34] http://www.doi.org/

Dadurch können Medienbrüche verhindert werden. Ein Archiv sollte gleichzeitig in der Lage sein alle gängigen Formate zu erkennen und darzustellen. So kann die Interoperabilität garantiert und ein Austausch von Information ermöglicht werden. Derzeit haben sich unter anderem folgende Formate durchgesetzt:

- Text: PDF, HTML, XML
- Bild: JPEG, TIFF
- Audio: MP3, WAV
- Video: AVI, MPEG
- Multimedia:
  Generell Komposition verschiedener Medien
  - Audiovisuell: MPEG, MXF

In diesem Zusammenhang sei hier das *Universelle Objektformat (UOF)*[35] aufgeführt, welches eine Paketstruktur einschließlich Metadaten beschreibt und somit als Archivformat und Austauschformat dient. Beliebige Arten von Dateistrukturen können von UOF aufgenommen werden. Das UOF transportiert sowohl inhaltliche als auch technische Metadaten. Im UOF wird mit einem Kern von Metadaten die komplette Migrationshistorie dokumentiert und alle Konvertierungen auflistet.

Im Bereich des Rundfunks wurde mit dem seit 2004 standardisierten *Material Exchange Format (MXF)* ein einheitliches und standardisiertes Dateiformat entwickelt unter der Zielsetzung, eine maximale Interoperabilität der verschiedenen Softwarekomponenten in der IT-basierten Fernsehproduktion zu erreichen. Das MXF ist ein Format für den Austausch von audiovisuellem Material und Metadaten

Öffentlich-rechtliche Rundfunkanstalten stellen Inhalte, wie sie auch ihren Archiven entstammen, zunehmend auch für mobile Endgeräte bereit. Daher sei hier auf die OMA[36]-Aktivitäten hinsichtlich einer interoperablen, offenen, weltweiten Standardisierung für mobilen Inhaltaustausch hingewiesen.

Metadatenformate

Hinsichtlich der Verfügbarkeit, Wiederauffindbarkeit, Plattformunabhängigkeit und Austauschbarkeit von Archivobjekten sollten Metadatenformate standardisiert werden. In diesem Kontext sind je nach Anwendungsbereich, hier Hochschul-Medienzentren und öffentlich-rechtliche Rundfunkanstalten, verschiedene Entwicklungen zu beobachten. Im Bereich der Hochschul-Medienzentren zählen unter anderem folgende zu den etablierten Standards:

Der Metadatenstandard *Dublin Core (DC)*[37] ist ein etablierter internationaler Standard (ISO 15836) zur Beschreibung von digitalen Objekten. Um Metadaten im Stil des *Dublin Core* bereitzustellen, wurde mit dem *Resource Description Framework (RDF)*[38] eine Infrastruktur entwickelt. RDF ist eine XML-basierte Sprache zur Spezifikation von Graphen (vgl. semantische Netzwerke). Der *Metadata Encoding and Transmission Standard (METS)*[39,40] ist ein von der *Digital Library Federation (DLF)* geförderter XML-basierter Standard zur Speicherung von digitalen Objekten mit ihren Meta- und Strukturdaten. Im Containerformat METS können verschiedene Strukturen (logische, physische etc.) abgebildet und verschiedene Metadatenstandards berücksichtigt werden. So bietet METS eine hohe Flexibilität und Konformität zum OAIS-Referenzmodell. Die PREMIS[41] (*Preservation Metadata*

---

[35] http://kopal.langzeitarchivierung.de/index_objektspezifikation.php.de

[36] http://www.openmobilealliance.org/

[37] http://dublincore.org/

[38] http://www.w3.org/RDF/

[39] http://www.loc.gov/standards/mets

[40] Library of Congress: METS: Überblick und Anleitung (Übersetzung: Angelika Menne-Haritz) Juli 2005.http://www.loc.gov/standards/mets/METSOverview.v2_de.html

[41] Data Dictionary for Preservation Metadata; Final Report of the PREMIS Working Group. 2005. Prepared on behalf of Preservation Metadata: Implementation Strategies (PREMIS), a working group jointly sponsored by OCLC and RLG. Available online at: http://www.oclc.org/research/projects/pmwg/premis-final.pdf

Implementation Strategies) als Fortführung der OCLC/RLG[42] *Preservation Metadata Working Group* sind als Rahmenkonzept für Metadaten etabliert, verbunden mit dem Ziel, implementierbare Metadaten mit größtmöglicher Anwendbarkeitsbreite für die Langzeitarchivierung festzulegen. Mit den Langzeitarchivierungsmetadaten für elektronische Ressourcen *(LMER)*[43] hat die Deutsche Nationalbibliothek einen universell ausgerichteten und XML-basierten Standard zur Generierung von technischen Metadaten entwickelt. Zur Generierung von Metadaten, insbesondere aber zur erweiterbaren Validierung von Formaten wurde *JSTOR/Harvard Object Validation Environment (JHOVE)*[44] etabliert und oftmals angewandt.

Im Zusammenhang mit Web Services sei hier auch der *Standard Web Ontology Language (OWL)*[45] erwähnt. OWL dient der formalen Beschreibung von Ontologien, so dass auch Software deren Bedeutung verarbeiten kann.

Im Kontext des Semantic Webs entwickelte sich der *Standard Topic Maps (ISO 2000b)*, der ähnlich wie RDF eine Beschreibungsform der Semantik von Objekten unter Zuhilfenahme von Netzstrukturen ist.

In Bezug auf Multimedia sind die von der *Moving Picture Experts Group* spezifizierten Standards MPEG-7 und MPEG-21 zu erwähnen. MPEG-7[46] (*Multimedia Content Description Interface Standard*) ist nicht darauf ausgelegt, den Inhalt selbst zu repräsentieren wie z.B. MPEG-4 für Multimedia. Mit dem Standard MPEG-7 wurde vielmehr ein Weg geschaffen, Information über den Inhalt bereitzustellen. Um die audiovisuellen Inhalte unabhängig von ihrer Speicherung, Darstellung, und Übertragung sowie unabhängig vom Medium oder den Technologien zu beschreiben, beinhaltet MPEG-7 einen umfassenden Satz von standardisierten Werkzeugen[47]. Auf diese Weise ist eine einfache, flexible und interoperable Lösung geschaffen für die Problematik des Erschließens und Katalogisierens (*Indexing*), Suchen und Abfragens (*Retrieval*) von Multimedia. Mit dem verwandten ISO/IEC-Standard MPEG-21[48,49] ist ein Multimedia-Rahmenwerk definiert, um die Prozesse zur Bereitstellung von Inhalten zu unterstützen. Der Standard MPEG-21 wurde spezifiziert, um den Austausch, den Zugriff und Handel, das Konsumieren und das Verändern von digitalen Multimediaangeboten auf effizientem, transparentem und interoperablem Weg zu ermöglichen.

Zusätzlich dazu wurden im Bereich der öffentlich-rechtlichen Rundfunkanstalten andere Standards entwickelt. Hierzu zählt das *Broadcast Metadata Exchange Format (BMF)*, ein einheitliches generisches Datenmodell für Metadaten in der Fernsehproduktion zur Sicherung der maximalen Interoperabilität zwischen allen Produktionsbereichen und Rundfunkanstalten.

Generell gibt es viele verschiedene Initiativen, was die Entwicklung von Metadatenstandards betrifft. In Bezug auf die Sicherstellung der Austauschbarkeit unter besonderer Berücksichtigung des Aspektes der Langzeit sollten zukünftige Entwicklungen in einem spezifischen Bereich darauf bedacht sein, die Metadatenstandards einheitlich zu gestalten.

6.1.2 Standardisierungen von Architektur und Infrastruktur

Mit dem OAIS-Referenzmodell ist bereits eine Standardisierung bzgl. der Architektur und Infrastruktur von digitalen Systemen erwirkt. Dieses Modell gibt den Rahmen für den Aufbau digitaler Langzeitarchivierungssysteme im Allgemeinen vor, lässt jedoch offen, wie eine Implementierung aussieht und berücksichtigt nicht den Einbezug der Sicherheit. Weitergehende Standardisierungsinitiativen sind dahingehend bereits von verschiedenen Konsortien getätigt bzw. werden ständig weiterentwickelt.

---

[42] http://www.oclc.org

[43] http://www.d-nb.de/standards/lmer/lmer.htm

[44] http://hul.harvard.edu/jhove/

[45] http://www.w3.org/2007/OWL/wiki/OWL_Working_Group

[46] http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

[47] http://xml.coverpages.org/mpeg7.html

[48] http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm

[49] http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=35366&ics1=35&ics2=040

Bezüglich des Standardisierungsbedarfs der Architektur und Infrastruktur wird an dieser Stelle auf die *Service-orientierten Architekturen* (SOA) verwiesen, die vermehrt zum Einsatz kommen. Erst kürzlich hat das Bundesamt für Sicherheit in der Informationstechnik eine Studie zu Sicherheit in Service-orientierten Architekturen[50] veröffentlicht. Mit diesem SOA Security Kompendium sollen Grundlagen und Handlungsrichtlinien für Entscheider und Entwickler bereitgestellt werden, sichere SOAs zu realisieren. Diese Studie zeigt, dass in diesem Bereich ein hoher Bedarf an Standardisierung vorherrscht. In der Studie werden Interoperabilitäts- und Kompatibilitätsaspekte benannt und laufende Standardisierungsprojekte vorgestellt. Es wird auf die Sicherheitsaspekte von SOAs und Web Services eingegangen wozu eine Bedrohungs- und Risikoanalyse der Standards gemacht wird und es wird eine Methodik zur Sicherheitsanalyse von Service-orientierten Architekturen vorgestellt.

SOAs zeichnen sich dadurch aus, dass sie Geschäftsprozess-nah sind und die Sicherheit wie Vertraulichkeit, Integrität und Authentizität für einzelne Service-Anfragen, Dienste und Prozesse einbeziehen können, wobei Sicherheitsanforderungen den einzelnen Service-Anfragen zugeordnet werden können. Geschäftsprozesse werden durch eine Aneinanderreihung von Aufrufen voneinander unabhängiger Services modelliert (*Composition of Services*). Funktionalitäten und Aufgaben werden als unabhängige, verkoppelte, austauschbare Dienste über standardisierte Schnittstellen von einem Service Provider angeboten. So kann eine hohe Flexibilität verbunden mit oftmals nicht unerheblichen Kostenersparnissen erreicht werden. Weiterhin können bestehende (Alt-)Systeme einfach eingebunden werden, Prozesse können kundengerecht und an sich ändernde Marktbedingungen und technische Weiterentwicklungen angepasst werden. Dies ist im Bereich der digitalen Langzeitarchivierung zur Bestandserhaltung von Bedeutung, denn SOAs bieten die Möglichkeit, Dienste oder Sicherheitsmechanismen kostengünstig und Aufwand minimierend einzubeziehen.

Ein SOA-Referenzmodell[51] wurde 2006 eingeführt von *OASIS Open Kompendium (*früher *SGML Open)*. In SOAs eingeschlossene Standardisierungen sind verbunden mit den bereitgestellten Web Services. Standards auch bzgl. der Sicherheit werden vom *World Wide Web Konsortium (W3C)*[52] und vom *OASIS*[53] *(Organization for the Advancement of Structured Information Standards)-Konsortium,* aber auch von anderen Arbeitsgruppen, Initiativen und Gremien entworfen und angeboten. Dazu zählen vorrangig XML (*eXtensible Markup Language*), SOAP (*Simple Object Access Protocoll*) und URI (*Uniform Resource Identifier*), sowie SOAP als Kommunikationsprotokoll (*W3C Recommendation)* zum Austausch von Daten zwischen Systemen. SOAP benutzt XML als Datenrepräsentationsmodell und http/TCP als Transportprotokoll.

Die vom W3C-Konsortium angebotenen Sicherheitsstandards sind u.a.:

- XML Encryption
- XML Signature
- XML Key Management Specification

Die vom OASIS-Konsortium angebotenen Sicherheitsstandards sind u.a.:

- Security Assertion Markup Language (SAML)
- WS-Security
  - WS-Trust
  - WS-Policy
  - WS-Federation

---

[50] Bundesamt für Sicherheit in der Informationstechnik - SOA Security Kompendium: *Sicherheit in Service-orientierten Architekturen*. 2008. URL: http://www.bsi.de/literat/studien/soa/index.htm

[51] OASIS Open: Reference Model for Service Oriented Architecture 1.0, OASIS Standard, 2006.

URL: http://www.oasis-open.org/specs/index.php#soa-rmv1.0 (Januar 2008)

[52] http://www.w3.org/

[53] http://www.oasis-open.org/home/index.php

Aktuell laufen innerhalb des OASIS-Konsortiums Initiativen zu:

- Biometric Identity Assurance Services (BIAS)
- Digital Signature Services eXtended (DSS-X)
- Enterprise Key Management Infrastructure (EKMI)

Für weiterführende Information dazu sei auf die Internetpräsentation von OASIS in Bezug auf Security[54] verwiesen.

Zu den Initiativen bzgl. verteilte Speichersysteme und Grid-Technologien wird in den folgenden Abschnitten 6.2 und 6.3 detailliert eingegangen.

Des Weiteren sei auf die vom Bundesministerium des Innern (BMI) erstellten und von der Koordinierungs- und Beratungsstelle der Bundesregierung für Informationstechnik in der Verwaltung (KBSt) herausgegebenen Standards und Architekturen für E-Government-Anwendungen (SAGA)[55] 3.0 verwiesen. Es beschreibt Standards, Verfahren und Methoden des Einsatzes der Informationstechnik in den Behörden und gibt Empfehlungen, insbesondere zur Gestaltung von E-Government-Angeboten der öffentlichen Verwaltung. „Mit der SAGA-Version 3.0 werden die Beschreibungen der Basiskomponenten, Infrastrukturkomponenten und Einer-für-Alle-Dienstleistungen (EfA-Dienstleistungen) aktualisiert und neue Themen wie "Beschreibungssprachen für Metadaten von Dateien", "Geodienste", "Langzeitarchivierung" und "Authentifizierung" werden eingebracht.“[56] Ähnlich wie bei SOA ist das Ziel von SAGA ein modernes, dienste-orientiertes Architekturkonzept. Der Bund will so sicherstellen, dass E-Government-Anwendungen interoperabel, plattformunabhängig und investitionssicher realisiert sind und ein reibungsloser Datenfluss im deutschen E-Government ermöglicht ist.

Ebenfalls vom Bundesministerium des Innern (BMI) erstellt und von der Koordinierungs- und Beratungsstelle der Bundesregierung für Informationstechnik in der Verwaltung (KBSt) herausgegeben ist das *DOMEA-Konzept* (Dokumentenmanagement und elektronische Archivierung im IT-gestützten Geschäftsgang). Das DOMEA-Organisationskonzept dient der sicheren Aufbewahrung integerer und authentischer und vollständiger Behördendokumente und beinhaltet dafür Standardisierungen.

Im Bereich der öffentlich-rechtlichen Rundfunkanstalten haben sich je nach Betriebsanforderungen zentrale Content Management Systeme (CMS) oder verteilte Systeme für das Management von Programminhalten in einer IT-basierten Produktionsumgebung bewährt.

6.1.3 Standardisierungen von Prozessen

Standardisierungen von Prozessen müssen zweigeteilt betrachtet werden, denn es muss zwischen

1. Prozesse innerhalb des Archivs – „intra“ und
2. Archivüberreifende Prozesse – „inter“

unterschieden werden. Während die Prozesse innerhalb des Archivs weitestgehend in dem OAIS-Referenzmodell abgebildet sind und damit hier Standardisierungen vorliegen, sind archivübergreifende Prozesse im OAIS-Referenzmodell nicht erfasst. Hier besteht ein dringender Handlungsbedarf bzgl. der Standardisierung.

6.1.4 Standardisierungen von Bestandserhaltungsstrategien

Generell muss die geeignete Bestandserhaltungsstrategie angewandt werden. Dies betrifft in erster Linie die Entscheidung zwischen Migration oder Emulation oder eine Verknüpfung beider. Hier ist ein Standardisierungsbedarf festzustellen, wie solche Strategien auszusehen haben, was sie in welchem Kontext beinhalten müssen, etc. Dies beruht auf Standardisierungen zu den anderen Punkten, wie Sicherheit, Formate, Hard- und Software, Metadaten, Netzwerke, Prozesse etc., denn dies sind alles Komponenten der Bestandserhaltungsstrategien und fließen dort ein.

---

[54] http://www.oasis-open.org/committees/tc_cat.php?cat=security

[55] http://www.kbst.bund.de/saga

[56] http://www.kbst.bund.de/saga

6.1.5 Standardisierungen in Bezug auf die Sicherheit

Für den Standardisierungsbedarf sei in Bezug auf die Sicherheitsanalyse digitaler Systeme, wie es auch digitale Langzeitarchivierungssysteme sind, auf die BSI-Standards [BSI08] und die IT-Grundschutz-Kataloge verwiesen, die Empfehlungen zu Methoden, Prozessen und Verfahren sowie Vorgehensweisen und Maßnahmen mit Bezug zur Informationssicherheit enthalten. Diese Standards sind dahingehend entwickelt worden, dass auf bewährte Methoden, Prozesse oder Verfahren zurückgegriffen werden kann, um sichere Nutzung von Informationstechnik zu erleichtern und um bewährte Herangehensweisen in ihrem Zusammenwirken darzustellen.

Die BSI-Standards [BSI08] umfassen:

- BSI-Standard 100-1: Managementsysteme für Informationssicherheit (ISMS)
- BSI-Standard 100-2: IT-Grundschutz-Vorgehensweise
- BSI-Standard 100-3: Risikoanalyse auf der Basis von IT-Grundschutz
- BSI-Standard 100-4: Notfall-Management (noch in interner Abstimmung)

BSI-Standard 100-1: Managementsysteme für Informationssicherheit (ISMS)

Der BSI-Standard 100-1: Managementsysteme für Informationssicherheit (ISMS) definiert allgemeine Anforderungen an ein ISMS und ist vollständig kompatibel zum ISO-Standard 27001 während er ebenso die Empfehlungen der ISO-Standards 13335 und 17799 berücksichtigt.

BSI-Standard 100-2: IT-Grundschutz-Vorgehensweise

Der BSI-Standard 100-2: IT-Grundschutz-Vorgehensweise beschreibt den Aufbau und die Anwendung von IT-Sicherheitsmanagement in der Praxis. Schwerpunkte sind die Benennung der Aufgaben des IT-Sicherheitsmanagements und des Aufbaus einer IT-Sicherheitsorganisation sowie die Erstellung eines IT-Sicherheitskonzepts in der Praxis, die Auswahl angemessener IT-Sicherheitsmaßnahmen und die Darstellung zu berücksichtigender Aspekte bei der Umsetzung des IT-Sicherheitskonzepts. Darüber hinaus wird in dem Standard beschrieben, wie IT-Sicherheit im laufenden Betrieb aufrechterhalten und verbessert werden kann. Mit diesem Standard werden die sehr allgemein gehaltenen Anforderungen der ISO-Standards 13335, 17799 und 27001 interpretiert, was Anwendern in der Praxis bei der Umsetzung mit vielen Hinweisen, Hintergrundinformationen und Beispielen helfen soll. Die IT-Grundschutz-Vorgehensweise zusammen mit den IT-Grundschutz-Katalogen erklären, was umgesetzt werden sollte und wie dies aussehen kann. Somit kann bei der Umsetzung allen Anforderungen der oben genannten ISO-Standards nachgekommen werden.

BSI-Standard 100-3: Risikoanalyse auf der Basis von IT-Grundschutz

Der BSI 100-3 zur Risikoanalyse wurde auf Grundlage der IT-Grundschutz-Kataloge erstellt, um eine ergänzende Risikoanalyse nahtlos an eine IT-Grundschutz-Analyse mit den IT-Grundschutz-Maßnahmen anschließen zu können. Möglichst viele Ergebnisse sollen so aus der IT-Grundschutz-Vorgehensweise wieder verwendet werden können, um wenig zusätzlichen Aufwand zu verursachen.

BSI-Standard 100-4 Notfall-Management

Der BSI-Standard 100-4: Notfall-Management befindet sich derzeit in der internen Abstimmung des BSI mit dem Ziel, eine stabile Version als Standard-Entwurf zur Diskussion zur Verfügung zu stellen. Mit dem BSI-Standard 100-4 soll ein Plan mit Anhaltspunkten erstellt werden, um bei Notfällen der verschiedensten Art richtig reagieren zu können.

IT-Grundschutz-Kataloge

Die BSI-Grundschutzkataloge [BSI06] dienen zur Erstellung von Sicherheitskonzepten. Sie werden an aktuelle Gegebenheiten angepasst und ständig erweitert. In den BSI-Grundschutz-Katalogen werden

Standard-Sicherheitsmaßnahmen für typische IT-Systeme empfohlen mit dem Ziel, durch geeignete Anwendung von organisatorischen, personellen, infrastrukturellen und technischen Standard-Sicherheitsmaßnahmen ein Sicherheitsniveau für IT-Systeme zu erreichen, das für den normalen Schutzbedarf angemessen und ausreichend ist und als Basis für hochschutzbedürftige IT-Systeme und -Anwendungen dienen kann.

Empfehlungen in Bezug auf die Sicherheit in Archivierung hat das BSI konkret in den IT-Grundschutz-Katalogen aufgeführt. Ausgehend von der Erhebung und Darstellung der Bausteine einer elektronischen Archivierung in B 1.12 Archivierung[57] wird die Gefährdungslage erhoben und Maßnahmen[58,59] werden vorgeschlagen.

ITSEC (Information Technology Security Evaluation Criteria) Kriterien60

Unter ITSEC-Kriterien sind Kriterien für die Bewertung der Sicherheit von Systemen der Informationstechnik (ITSEC) aufgeführt, die bereits 1991 von der Europäischen Kommission veröffentlicht wurden. ITSEC ist ein europäischer Standard für die Bewertung und Zertifizierung von Software und IT-Systemen in Hinblick auf ihre Funktionalität und Vertrauenswürdigkeit bezüglich der Daten- und Computersicherheit.

CC-Common Criteria (ISO/IEC 15408)

Die Common Criteria sind gemeinsame Kriterien für die Prüfung und Bewertung der Sicherheit von Informationstechnik (*Common Criteria for Information Technology Security Evaluation, CC*) und dienen der Bewertung der Sicherheitseigenschaften von IT-Systemen und Produkten. Die CC sind als internationaler Standard ISO/IEC 15408 anerkannt. Die CC sind eine Weiterentwicklung und Zusammenführung der europäischen Kriterien für die Bewertung der Sicherheit von Systemen der Informationstechnik (ITSEC), des Orange-Book (TCSEC) der USA und der kanadischen Kriterien (CTCPEC). Für die Sicherheitsevaluation von digitalen Langzeitarchivierungssystemen sind die CC heranzuziehen und hier als Standard aufgeführt.

Digitale Signaturen

Generell ist der Einsatz von digitalen Signaturen in digitalen Langzeitarchivierungssystemen empfohlen, da sie grundsätzlich der Sicherheit dienen und in diesem Bereich Aktivitäten zu Standardisierungen erfolgreich laufen. TeleTrust Deutschland e.V.[61], Verein zur Förderung der Vertrauenswürdigkeit in der Informations- und Kommunikationstechnik beschäftigt sich seit Jahren mit dem standardisierten Aufbau der Infrastrukturen für qualifizierte elektronische Signaturen in Deutschland und dem europäischen Ausland. In Zusammenarbeit mit der T7-Arbeitsgemeinschaft[62] der Trustcenterbetreiber wurde die Interoperabilitätsspezifikation ISIS-MTT[63] entwickelt. Der ISIS-MTT Standard definiert Spezifikationen von Sicherheitsfunktionalitäten für elektronische Signaturen und sichere E-Mails sowie für das IT-Sicherheitsmanagement, was sich auf Public Key Infrastrukturen stützt. Mit dem ISIS-MTT Standard ist die Interoperabilität geschaffen.

Die Grundlagen für digitale Signaturen sind im *Signaturgesetzt SigG* bestimmt. Darauf aufbauend legt das BSI fest, welche Verfahren und Schlüssellängen einzusetzen sind, wenn man qualifizierte Signaturen anwenden will. So liegt derzeit eine aktuelle Änderung vom BSI vor, dass ab Januar 2008 bis voraussichtlich 2011 die Schlüssellängen für qualifizierte Signaturen 2048 Bit betragen müssen und eine Länge von 1024 Bit nicht mehr als sicher gilt. Wenn keine qualifizierten Signaturen ein-

[57] http://www.bsi.de/gshb/deutsch/baust/b01012.htm

[58] http://www.bsi.de/gshb/deutsch/m/m01059.htm

[59] http://www.bsi.de/gshb/deutsch/m/m01060.htm

[60] http://www.bsi.bund.de/zertifiz/itkrit/itsec.htm

[61] http://www.teletrust.de

[62] http://www.t7-isis.org

[63] http://www.isis-mtt.org

gesetzt werden, muss sich weder an vorgegebene Verfahren noch an Schlüssellängen gehalten werden. Solche Signaturen haben aber im Gegensatz zu den qualifizierten Signaturen keine Rechtsgültigkeit.

Generell gibt es PKCS-Standards als Formatstandards für digitale Signaturen. Der derzeit gängigste Standard ist der PKCS-7-Standard, wird aber zunehmend vom PKCS-11-Standard verdrängt. Diese Standards sind deswegen festgelegt worden, damit ein anderes Programm die Signatur auch lesen kann.

Im Zusammenhang mit digitalen Signaturen sei hier beispielsweise auf ArchiSig[64], Beweiskräftige und sichere Langzeitarchivierung digital signierter Dokumente verwiesen.

Zusammenfassung

Zusammenfassend ist festzuhalten, dass bereits eine Vielzahl von Aktivitäten in Bezug auf Standardisierungen läuft, die eine *vertrauenswürdige und abgesicherte digitale Langzeitarchivierung* betreffen und dort bereits angewendet werden und potentiell angewendet werden können. Es wurde aufgezeigt, dass gerade was den Aspekt der Sicherheit betrifft, bereits Vorarbeit geleistet wurde, um Sicherheitsmechanismen in bestehende Systeme der digitalen Langzeitarchivierung zu integrieren. Der Handlungsbedarf gestaltet sich nun dahingehend, Sicherheitsmechanismen in die Funktionalitäten und Prozesse der digitalen Langzeitarchivierung standardisiert einzubeziehen. Dies zum einen innerhalb eines Archivs, um den Bestand nachhaltig zu erhalten und auch in Zukunft nachhaltig zu sichern, aber zukünftig auch um eine sichere Kommunikation nach Außen zu gewährleisten, ohne dabei den Archivbestand zu gefährden.

6.2 Initiierung und Engagement von Aktivitäten des Kompetenznetzwerks als Ganzes für hochgradig verteilte Speicherlösungen mit ad-hoc Ressourcenbedarf

Bedingt durch die Auswirkungen moderner Informations- und Kommunikationstechnologie sind die Produktion, die Verbreitung, die Nutzung und die Bewahrung von Medien bereits seit geraumer Zeit tief greifenden Veränderungen unterworfen. Schon heute ist diese Entwicklung gekennzeichnet durch eine Beschleunigung der Produktions- und Wiederverwendungszyklen sowie eine rasant zunehmende Fülle und Vielfalt der bereitstehenden und umgesetzten Medieninhalte. Der Wandel erfasst ebenfalls die Arbeitsweise, neue Formen der Zusammenarbeit über die Grenzen von Organisationen und technische Plattformen hinweg bilden sich heraus. Einher geht ein allgemeiner Trend zur zunehmenden Verflechtung der Akteure bis hin zur globalen Ebene.

Grid-Computing bezeichnet die integrierte, gemeinschaftliche Verwendung von verteilten Ressourcen. Der Ansatz erschließt sowohl Daten als auch Rechner, Speicher, Instrumente, Anwendungen, Sensoren und Dienste allgemein. Über die klassischen Methoden der Ressourcenverteilung und –verwaltung hinaus erlaubt Grid-Technologie den Aufbau flexibler Umgebungen mit einer transparenten, bedarfsgerechten Ressourcen-Bereitstellung und –Nutzung. Virtualisierung leistet hierbei die Abstraktion von der tatsächlichen Beschaffenheit und räumlichen Verteilung der einzelnen Ressourcen. Nutzer müssen sich in Folge mit der Heterogenität der Ressourcen nicht weiter auseinandersetzen und können sich auf deren Leistungsmerkmale konzentrieren.[65]

Anwendungsbereiche für die Grid-Technologie in Wissenschaft und Privatwirtschaft zeichnen sich durch eine hohe Dynamik, ein hohes Datenaufkommen, einen hohen Recheneinsatz und einen Bedarf nach Langzeitarchivierung zwecks Nachweis, Kulturgutbewahrung oder Zweitverwertung aus. Um-

[64] http://www.archisig.de/

[65] Wegbereiter für das Grid-Computing war die Leistungssteigerung der Kommunikationsnetze. Die darauf aufbauende Infrastruktur des Grid leistet die Koordination der eingebundenen Ressourcen und mit deren Nutzung verbundene Verwaltungsaufgaben wie beispielsweise Abrechnung und Zugriffskontrolle. Standardisierte, offene Protokolle und Schnittstellen des Grid stellen Funktionen für Authentifizierung, Autorisierung, Ressourcen-Ermittlung, Ressourcen-Zugriff und Weitere bereit. Für die Ressourcennutzung kann zudem eine Dienstgüte wie beispielsweise Antwortzeit, Durchsatz, Erreichbarkeit oder Sicherheit vereinbart werden.

gekehrt könnte diese Technologie ein großes Potenzial für die Implementierung von Langzeitarchivsystemen in sich bergen.

Das deutsche Kompetenznetzwerk zur digitalen Langzeitarchivierung nestor hat die Tragweite des Grid für die Langzeitarchivierung erkannt und einen entsprechenden inhaltlichen Themenschwerpunkt in seine Arbeit aufgenommen. Die gebildete Arbeitsgruppe Grid[66] verfolgt als Ziel, die *Chancen und Risiken* dieses neuen Gebietes zu skizzieren und eine *Roadmap* zu seiner Erschließung als auch zur Gridifizierung von Archivsystemen aufzustellen. Drei parallel zu dieser Expertise laufende Studien untersuchen dieses Gebiet unter den Gesichtspunkten Rohdaten, Synergien respektive Standards:

- "Anforderungen von e-Science und Grid-Technologie an die Archivierung wissenschaftlicher Rohdaten", GeoForschungszentrum Potsdam, Dr. Jens Klump [Klu08]
- "Synergiepotenziale zwischen GRID- und e-Science-Technologien für die Langzeitarchivierung", FernUniversität in Hagen, Prof. Dr. Wolfram Schiffmann [Sch08]
- "Standards und Standardisierung im Kontext von Grid-Technologien und Langzeitarchivierung", Universität der Bundeswehr München, Prof. Dr. Uwe M. Borghoff [Bor08]

Diese vorliegende Expertise widmet sich der grundsätzlichen Fragestellung, inwieweit Sicherheitstechnologien zu einer vertrauenswürdigen und abgesicherten Langzeitarchivierung multimedialer Inhalte beitragen können. Die Betrachtung gründet auf einer Erhebung der in der Praxis gegenwärtig eingesetzten oder sich im Aufbau befindlichen technischen Infrastrukturen in den zwei exemplarischen Szenarien öffentlich-rechtliche Rundfunkarchive sowie Hochschul-Medienzentren. Im Ergebnis zeigt deren Untersuchung Herangehensweisen für die Integration von Sicherheitstechnologien auf und ermittelt weiteren Handlungsbedarf. Im Sinne einer zielgerichteten Gestaltung der zukünftigen Entwicklung bezieht diese Expertise im Folgenden zusätzlich das Grid als eine sich abzeichnende fortschrittliche Technologie mit Relevanz für die Langzeitarchivierung in ihre Betrachtung mit ein. Ergänzend und im Zusammenspiel mit den anderen drei Studien skizziert sie durch die Grid-Technologie mögliche Änderungen des Archivumfelds sowie Potenziale eines Einsatzes von Grid-Technologie zu Archivaufgaben, schätzt ihre Auswirkungen auf die vertrauliche und abgesicherte Langzeitbewahrung ab und gibt einen Ausblick auf den Bedarf von sicherheitstechnischen Maßnahmen.

Generell besteht die Notwendigkeit von Maßnahmen zur Etablierung von Sicherheit in Grid-Umgebungen, um die Unverfälschtheit der verteilt verwalteten Inhalte zu gewährleisten. Maßnahmen stützen sich auf Gefährdungen, wie sie für Archive im BSI aufgelistet sind. So müssen die Sicherheitsaspekte Verfügbarkeit, Integrität, Authentizität, Vertraulichkeit und Nachweisbarkeit immer neu bewertet und die Gefahren erhoben werden. Dies gilt universell für das digitale Langzeitarchiv insgesamt, aber auch für alle Prozesse, von *Ingest, Access, Archival Storage, Data Management* bis *Preservation Planning*. Sicherheitsmaßnahmen geschehen dabei auf den Ebenen IT-Systeme und Netze, Anwendungen, Infrastruktur und übergeordnete Aspekte (IT-Grundschutz-Kataloge). Je nach Bestand und Anwendungsgebiet des Archivs müssen beispielsweise Urheberrechte geschützt, der Jugendschutz berücksichtigt, Nutzungsrechte beachtet und eingebunden, Zugangssicherheit zum Archiv (sichere Authentifizierung) hergestellt und fehlerfreie Übertragung sichergestellt werden. Darüber hinaus müssen vor allem auch Maßnahmen getroffen werden, um ein digitales Langzeitarchiv mit seinem Bestand gegen Programme mit Schadensfunktion und unautorisierte Zugriffe sowie Manipulationen zu sichern. Dies wird in Zukunft besonders wichtig, wenn Archive nach außen kommunizieren und Archivobjekte austauschen oder miteinander verknüpfen.

In einer Grid-Umgebung lassen sich Sicherheitsmechanismen wie digitale Signaturen und Zertifikate von Trustcentern als Dienste einbinden. Dies ist in der dynamischen Umgebung von digitalen Langzeitarchiven nicht ausreichend. Zusätzliche Sicherheits-relevante Vorkehrungen müssen getroffen werden, um die Vertrauenswürdigkeit, Integrität und Authentizität zu gewährleisten. So sollten beispielsweise Mechanismen angewandt werden, wie DRM-Systeme, Verschlüsselungstechniken oder digitale Wasserzeichen, mit denen Nutzungsrechte erzwungen, Urheberrechte eingehalten, vertrauenswürdige Dienstleister auf technischem Weg identifiziert und autorisiert, eine sichere Kommunikation und Infrastruktur etabliert sowie integere und authentische Ressourcen nachgewiesen werden können.

[66] Siehe http://www.langzeitarchivierung.de, Arbeitsgruppen

Sicherheitsmechanismen werden an dieser Stelle nicht detailliert im Einzelnen aufgeführt und erklärt, es wird stattdessen auf die Fachliteratur [Dit00] sowie [Eck03] verwiesen.

Mit der Einbindung von den in der IT-Sicherheit bereitgestellten Sicherheitsmechanismen können digitale Langzeitarchivierungssysteme vertrauenswürdig und sicher gestaltet werden. Wie sich dies auf das Vertrauen seitens des Benutzers auswirkt, ist beispielsweise in [OeDi06] erörtert.

Sicherheit muss auf allen Ebenen eines digitalen Langzeitarchivs gewährleistet sein:

- Komponenten und Ressourcen (Hard- und Software) wie Speichermedien und Darstellungsanwendungen, Datenbanken, Clients
- Metadaten
- Digitales Objekt (Inhalt)
- Netzwerkverbindungen
- Prozesse (*Ingest*, *Access*, *Archival Storage*, *Data Management*, *Preservation Planning*)

Sicherheitsmechanismen sollten entsprechend angewandt werden, um gegebene Sicherheitsanforderungen bzgl. der Aspekte Verfügbarkeit, Integrität, Authentizität, Vertraulichkeit und Nachweisbarkeit zu erbringen. Für *Ingest* sollten im Sinne einer vertrauenswürdigen und abgesicherten Langzeitarchivierung digitaler Inhalte Mechanismen zur Verfügung stehen, mit denen das durch den Produzenten vorproduzierte Datenmaterial auf Integrität und Authentizität überprüft und validiert werden kann. Hier können beispielsweise Sicherheitsmechanismen, welche aus dem Bereich der digitalen Medienforensik stammen, Anwendung finden. Für *Access* sollten Sicherheitsmechanismen zur Identifizierung des zugreifenden Konsumenten bereitstehen. Die hier benannten Sicherheitsmaßnahmen sind beispielhafte Vorschläge, die keinesfalls vollständig sind.

In Hinblick auf eine *vertrauenswürdige und abgesicherte* Langzeitarchivierung bedarf es zusätzlichen Engagements von Seiten des Kompetenznetzwerkes. Für eine Einschätzung und Sicherstellung der Vertrauenswürdigkeit von digitalen Informationsobjekten ist deren gesamter Lebenszyklus von der Entstehung bis zur Archivierung und Wiederverwendung von Relevanz. Weil dieser Lebenszyklus zunehmend in verteilten Umgebungen stattfinden wird, kann eine Sicherheitsbetrachtung nicht an den Grenzen des Langzeitarchivs enden. Dies gilt umso mehr, je stärker die Langzeitarchivierung operativ in Grids eingebunden wird und zur Erbringung seiner Aufgaben selbst Ressourcen dieser verteilten Infrastruktur nutzt. Folgende Aktivitäten des Kompetenznetzwerkes können zu einer vertrauenswürdigen und abgesicherten Langzeitarchivierung im Grid beitragen:

- Zusammenarbeit mit Gremien der Grid-Entwicklung zur Einbringung von Anforderungen der vertrauenswürdigen und abgesicherten Langzeitarchivierung in die weitere Spezifikation.
- Untersuchung der Leistungsfähigkeit im Grid implementierter Sicherheitsmechanismen bezüglich Verfügbarkeit, Integrität, Authentizität und Vertraulichkeit [Klu08].
- Prüfung heutiger Authentifizierungs- und Autorisierungsverfahren auf ihre Eignung für eine Verwendung auf lange Zeit und die Möglichkeiten ihrer sicheren Übertragung auf neue Verfahren [Klu08].

6.3 Potentiale für den Einsatz von Grid- und anderen Virtualisierungstechnologien

In Bezug auf Multimediaarchive lassen sich zwei potenzielle Einsatzfelder für Grid-Technologien (Abbildung 38) ausmachen:

1. Bereitstellung virtualisierter Langzeitarchivierungsdienste in einer Grid-Umgebung; Das Grid nutzt das Archiv.
2. Nutzung virtualisierter Rechen- und Speicherkapazitäten einer Grid-Umgebung für Aufgaben des Langzeitarchivs; Das Archiv nutzt das Grid.

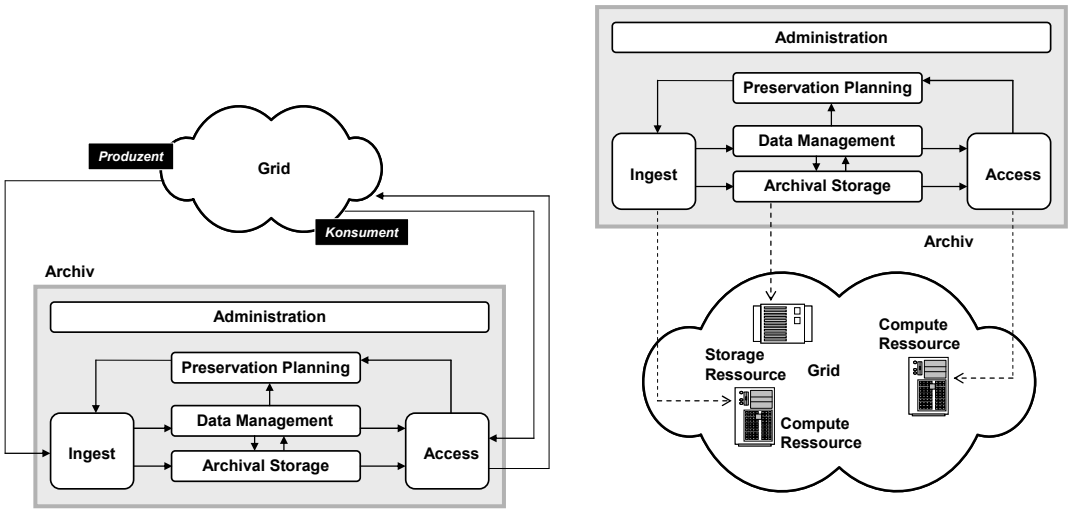

Abbildung 38: Potentiale für den Einsatz von Grid- und anderen Virtualisierungstechnologien für die Langzeitarchivierung multimedialer Inhalte: Bereitstellung virtualisierter Langzeitarchivierungsdienste in einer Grid-Umgebung (links) und Nutzung virtualisierter Rechen- und Speicherkapazitäten einer Grid-Umgebung für Aufgaben des Langzeitarchivs (rechts).

An dieser Stelle beschränkt sich die Betrachtung auf technische Fragen in Bezug auf die Archivierung großer und komplexer Mediendatenbestände. Fragen der Organisation von Arbeitsabläufen in der Produktion und Nutzung von Medieninhalten und Fragen notwendiger Rahmenbedingungen für eine erfolgreiche Langzeitarchivierung digitaler Daten sind nicht Bestandteil.

6.3.1 Bereitstellung virtualisierter Langzeitarchivierungsdienste in einer Grid-Umgebung

Die voran geschrittene Digitalisierung hat die wesentlichen Voraussetzungen für den einfachen und schnellen Zugriff auf archivierte Multimedia-Datenbestände geschaffen.

Die Grid-Technologie ermöglicht eine einheitliche Vernetzung und Bereitstellung von entfernten Ressourcen und Diensten. Mittels standardisierter Schnittstellen für *Ingest* und *Access* können heterogene Archive als spezifische Grid-Services in eine Grid-Umgebung eingegliedert werden (Abbildung 38 links). So erschließen sich deren multimedialen Daten weiteren Diensten einheitlich für eine verteilte Verarbeitung, Aufbereitung als auch für eine Weiterverwertung und werden hochverfügbar [Sch08].

Die Archive der öffentlich-rechtlichen Rundfunkanstalten haben einen Doppelauftrag: Einerseits Archivierung und Dokumentation und andererseits Nutzbarmachung für Wiederverwertung. Das *Programmvermögen* der Anstalten hat einen hohen kulturellen als auch wirtschaftlichen Wert. Es wird erwartet, dass sich die ohnehin bereits hohe Wiederverwertungsquote bei einer konsequenten Weiterentwicklung zu einem direkten, öffentlichen Zugriff auf die Inhalte noch steigern lässt. Auch bei Beschränkung des Zugriffs auf ausgewählte Benutzergruppen wie beispielsweise Wissenschaftler oder Zweitverwerter wären Vorteile zu erkennen. Die Rundfunkarchive könnten aus der Vergabe von Nutzungsrechten ihr Programmvermögen wirtschaftlich verwerten. Für Wissenschaftler als Konsumenten vereinfachte sich der Zugang zu Quellmaterial für ihre Forschungsarbeit.

Im Wissenschaftsprozess werden auch vergangene respektive historische Fakten immer wieder neu bewertet werden. Die Langzeitarchivierung digitaler Ressourcen ist daher eine wesentliche Bedingung für die Konkurrenzfähigkeit sowohl des Bildungs- und Wissenschaftssystems als auch der Wirtschaft. Aktuelle Bemühungen in der Organisation der wissenschaftlichen Arbeit gehen dahin, zusätzlich zu der traditionellen Publikation die im Forschungsprozess gewonnenen und zu der Erkenntnis führenden Rohdaten ebenfalls aufzubewahren und bereitzustellen. In der modernen Wissenschaft kommen immer aufwändigere Detektoren und Simulationen zum Einsatz. Die von ihnen erzeugten Experimentaldaten sprengen mit ihrem Volumen vorher bekannte Grenzen und sind in ihren Formaten vielfältig. Auch Audio- und Videoaufzeichnungen von Experimenten und Beobachtungen sowie Visualisierungen, also typisches Multimediamaterial, sind regelmäßig Bestandteil des wissenschaftlichen Prozesses.

Eine Öffnung und Bereitstellung der Multimediaarchive ist Voraussetzung für die umfassende Verwertung der Sammlungen. Die bislang oftmals unabhängige heterogene Entwicklung der Archive,

unter anderem auch bei öffentlich-rechtlichen Rundfunkarchiven, ist ein Hemmnis. Auch wenn Medienarchive über Informationsnetze erreichbar sind lassen sie sich in aller Regel nicht einheitlich ansprechen. Standardisierte Grid-Schnittstellen ermöglichten auch Institutions-übergreifende Recherchen.

Mit der Erschließung der Multimedia-Archive erhalten Drittanbieter die Gelegenheit, auf Basis der Kollektionen Multimedia-Dienstleistungen sowie –Anwendungen zu entwickeln und zu vermarkten. Exemplarisch angeführt werden können die Analyse von Multimedia-Inhalten, fachspezifische Recherchen, Benachrichtigungsdienste oder die redaktionelle und technische Aufbereitung für neuartige Verbreitungswege.

Gleichzeitig ist diese Entwicklungsperspektive verbunden mit einem verstärkten Bedrohungspotential, da das Archiv nicht länger als überwiegend abgeschlossene Einheit zu betrachten ist. Vielmehr öffnet sich das Archiv gegenüber Produzenten und Konsumenten als auch gegenüber anderen Archiven und steht mit ihnen in interaktiver Kommunikation. Hier bedarf es eines intensiven und gezielten Einsatzes von Sicherheitsmechanismen, um die Vertrauenswürdigkeit zu gewährleisten, die Integrität und Authentizität des Bestandes zu sichern und die kommunizierenden Parteien zu autorisieren.

6.3.2 Nutzung virtualisierter Rechen- und Speicherkapazitäten einer Grid-Umgebung für Aufgaben des Langzeitarchivs

Potenzial für einen Einsatz virtualisierter Rechen- und Speicherkapazitäten ist insbesondere für das *Archival Storage*, den *Ingest* und den *Access* absehbar, bislang nicht jedoch für die Funktionsbereiche *Preservation Planning, Data Management* und *Administration* des Archivs.

Archival Storage

Mit der Digitalisierung der multimedialen Inhalte reduziert sich deren Repräsentation auf Daten. Als Bitfolge unterscheiden sie sich nicht von den Daten jeder anderen Computeranwendung. Mediendateien zu Archivzwecken zu speichern reduziert sich darauf, Dateien auf Datenträgern zu lagern und zuverlässig wieder auszulesen. Zugleich haben Archive aufgrund ihres Erhaltungsauftrages einen stetig wachsenden Speicherbedarf.

Die Grid-Technologie ermöglicht eine einheitliche Vernetzung und Bereitstellung von entfernten Ressourcen und Diensten. Auf diese Weise können Grid-Services zur Speicherung und zuverlässigen Bereitstellung von Daten aufgebaut und angeboten werden (Abbildung 38 rechts). Medienarchive erhalten die Möglichkeit, ihre Speicherung von Mediendaten, das *Archival Storage*, an diese Dienste abzugeben [Sch08] und sich auf ihre spezifischen Aufgaben *Ingest, Access, Administration* und *Preservation Planning* zu konzentrieren.

Die immerwährende Nachfrage der Archive nach mehr Speicher wird durch ein atemberaubendes Wachstum der Kapazitäten der verfügbaren Speichertechnologien bedient. Doch während aufgrund technischer Entwicklungen die Kosten pro Speicherplatzeinheit sich im rasanten Fall befinden, steigen dem entgegen die Kosten für die administrative Verwaltung von installierten Speichersystemen. „Analysten […] schätzen das Verhältnis von Hardware- zu Managementkosten mittlerweile auf rund 1:3.“ [Chr06] Die Aufwendungen für die Verwaltung der Speichersysteme, das Datenmanagement sowie die Schulung des Personals und den Support werden zu einem bestimmenden Kostentreiber für die langfristige Bewahrung der Archivalien. Die hohen Verwaltungskosten für den Betrieb von Speichersystemen sind eine starke Motivation, diese Aufgabe auszulagern.

Drei aktuelle Fallbeispiele aus Norwegen, den Niederlanden und Großbritannien illustrieren die latente Neigung, dass die Digitalisierung der Archivalien zu einer *Trennung von Speicherung und Archiv* führt [Wri07]. In den benannten Fällen übernahmen externe Dienstleister die Speicherung und Bereitstellung der Inhalte von Rundfunkarchiven. Auch in Deutschland haben einige private Sendeanstalten die Speicherung der Archivinhalte an Dienstleister ausgelagert oder beabsichtigen, dieses in der nahen Zukunft zu tun. Diese Organisation wird von den Archivaren als nicht problematisch betrachtet, sondern vielmehr als Erleichterung empfunden. Voraussetzung ist allerdings, dass die Speicherung stabil und lang anhaltend ist, der Speicher groß und schnell genug zugreifbar ist, und dass die ausgelesene Datei nutzbar ist [Wri07].

Jedoch legen Archivare an die Datenhaltung andere Qualitätskriterien an als die von der Speicherindustrie verwendeten [Wri07]. Während für die Langzeitarchivierung Stabilität (*Persistence*) von zentraler Bedeutung ist, benutzt die Speicherindustrie die Begriffe Fehlerrate des Ausleseprozesses, Fehlerrate des Gerätes und Lebenserwartung des Datenträgers. Für das Archivwesen wesentliche *Bemessungsgrößen einer Speicherstrategie* zur Planung der Bestandserhaltung sind hingegen:

- Wie viel Inhalt wird in einem gegebenen Zeitraum verloren gehen?
- Wie verhält sich die statistische Verteilung des möglichen Verlustes?
- Wie verändern sich die Wahrscheinlichkeiten über die Speicherdauer?
- Wie wirken sich die Wahrscheinlichkeiten auf die Kosten aus?

In den durchgeführten Auslagerungsprojekten war die Auswertung von Statistiken notwendig, um diese Qualitätskriterien bestimmen zu können. Für den Aufbau von Grid-Services zur Langzeitspeicherung von Daten wären diese Eigenschaften essenzielle Angaben in der Beschreibung eines solchen Dienstes.

Auf welche Weise eine Speicherdienstleistung erbracht wird, ist aus Sicht des Archivs letztendlich unwichtig, solange die angesetzten Qualitätskriterien erfüllt werden. Unter diesem Gesichtspunkt ist die Kapselung des *Archival Storage* in strikt spezifizierten Diensten ideal.

Der Erhaltungsauftrag sorgt für einen stetig wachsenden Speicherbedarf der Archive. Im Bereich der Rundfunkarchive bestimmt sich dieser anhand der fortwährenden Aufzeichnung der Sendeströme (Vorwärtsdigitalisierung) und die nachträgliche Überführung auf Film- und Videoband vorbestehender Archivalien in das digitale Archiv (Rückwärtsdigitalisierung). Die adäquate Planung mittelfristig benötigter Speicherkapazitäten ist grundlegender Bestandteil der Konzeption einer Archivierungslösung. Absehbare Entwicklungen wie die Ausspielung auf neue Verbreitungswege oder hochaufgelöstes Video (*High Definition, HD*) lassen ein drastisches Ansteigen des Volumens neu zu archivierender Inhalte erwarten. Skalierbarkeit ist ein bedeutendes Leistungsmerkmal der in Archiven zum Einsatz kommenden Speichersysteme. Dynamisch vermittelte und gebundene Grid-Services zur Speicherung und zuverlässigen Bereitstellung von Daten haben das Potenzial, kurzfristige und vorübergehende Spitzen im Speicherplatzbedarf aufzunehmen.

Stellten Archive ihre Datenhaltung vollständig auf externe Grid-Services um, so bedeutete die dynamische Inanspruchnahme zusätzlicher Speicherkapazitäten keinen Ausnahmefall, sondern wäre vielmehr ein regulärer Arbeitsschritt in der fortlaufenden Bestandserweiterung. Für das Archiv entfiele der Aufbau einer eigenen Infrastruktur für das *Archival Storage*. Die Vorteile eines solchen Modells sind die geringere Investitionstiefe und die rein bedarfsorientierte Abrechnung – das Archiv zahlt lediglich für den Speicher, den es auch tatsächlich nutzt.

Die Aufbewahrung mehrerer Kopien verteilt auf unterschiedliche Instanzen ist eine mögliche Strategie, um Datenverlust vorzubeugen. Mit seiner verteilten Architektur wohnt dem Grid diese Eigenschaft bereits inne. Die Beauftragung verschiedener Grid-Services mit der Speicherung von Archivalien hat das Potenzial, diese Archivalien vor einer Zerstörung durch lokal eingrenzbare Havarien zu schützen. Voraussetzung ist eine gezielte Streuung mehrfacher Kopien einer Datei, entgegen einem Grundprinzip der Virtualisierung.

Mit heutigem Stand ist für den Bereich der öffentlich-rechtlichen Rundfunkanstalten in Deutschland eine Speicherung und Archivierung in über verschiedene Standorte verteilte Grids im Sinne von Grid-Computing technisch noch nicht absehbar. Im Europäischen Ausland nutzt eine Rundfunkanstalt in einem Pilotprojekt Grid-Technologien, um Multimediadaten über mehrere eigene Standorte zu spiegeln.

Es sind allerdings erste Plattenspeichersysteme verfügbar, die intern als Grids aufgebaut sind. Obgleich ihrer Namensverwandtschaft sind Grid-Storage und Grid-Computing grundsätzlich voneinander unabhängige Technologien [Str04].

Bei einer klassischen Beauftragung eines externen Dienstleisters für die Datenhaltung sind Auftraggeber und Dienstleister gegenseitig bekannt. Bei Nutzung von Grid-Services zur Speicherung und zuverlässigen Bereitstellung von Archivdateien steht das Archiv einer mehr oder minder anonymen Infrastruktur gegenüber, welcher es seine oft wertvollen Inhalte anvertraut. Eingespeiste Daten bedürfen eines Schutzes. Für eine vertrauenswürdige, abgesicherte Langzeitarchivierung sind Nachweis-

barkeit, Integrität, Authentizität, Verfügbarkeit und Vertraulichkeit sicherzustellen. Weiterhin sind Urheber- und Nutzungsrechte zu wahren. Verschlüsselungsverfahren und digitale Wasserzeichen sind mögliche Ansätze, um einen derartigen Schutz zu gewährleisten.

Ingest und Access

Mit der Digitalisierung der Medieninhalte und unter Verwendung moderner Kommunikationstechnologie ist, ohne die Bindung an einen physischen Datenträger, die Weitergabe auch umfangreicher multimedialer Daten zwischen den Stationen eines Medien-Produktionsprozesses bereits handhabbar.

Die Grid-Technologie ermöglicht eine einheitliche Vernetzung und Bereitstellung von entfernten Ressourcen und Diensten. Auf dieser Basis können ebenfalls *Grid-Services* zur Bearbeitung, Analyse, Indizierung, Suche, Kontextualisierung und Verbreitung multimedialer Daten entstehen (Abbildung 38 rechts). Medienarchive erhalten die Möglichkeit, Teilaufgaben von *Ingest* und *Access* an derartige Dienste abzugeben. Das Langzeitarchiv erbringt in diesem Fall Mehrwert anhand der in seiner Hoheit verbleibenden Koordination der ausgelagerten Funktionen mittels *Administration* und *Data Management* sowie die strategische Planung, dem *Preservation Planning*.

Im Bereich des Rundfunks ist mit der vernetzten Produktionsumgebung eine verteilte und vernetzte Arbeitsweise innerhalb der Rundfunkanstalten zum heutigen Tag Realität. In der Zusammenarbeit von Medienunternehmen ist ein Trend hin zu eigenständigen, hoch spezialisierten Einheiten zu erkennen, welche Leistungen innerhalb der Wertschöpfungskette erbringen. Bei der auf Informationstechnik gestützten Produktionsweise wird sich im Rundfunk eine dezentrale Bearbeitung der Beiträge durchsetzen. Die Produktionsleistungen werden dort erbracht, wo die höchste Kompetenz am günstigsten angeboten wird und die erreichte Qualität und Bearbeitungszeit die Vorgaben erfüllen. [Sau07] Die Grid-Technologie bietet das Potential, dedizierte Multimedia-Dienstleistungen und Anwendungen über ihre Infrastruktur anzubieten.

Die großen Datenmengen gestalten eine entfernte Exploration multimedialer Archive schwierig. Der Selektion und Suche nach spezifischen Inhalten kommt ein hoher Stellenwert zu, wenn sie nicht gar unabdingbare Voraussetzung für das gezielte Auffinden relevanter Inhalte sind.

Für die Analyse von Audio-, Sprach-, Bild- und Videodaten zu Zwecken ihrer automatisierten Indizierung steht derzeit bereits eine Vielzahl von Algorithmen bereit, beziehungsweise befindet sich in der Entwicklung, unter anderem:

- Zeitliche Segmentierung von Videodokumenten
- Erkennung von Text
- Erkennung von Gesichtern
- Erkennung von Sprechern
- Erkennung von Sprache
- Erkennung von Kamerabewegungen
- Extraktion von Schlüsselbildern
- Extraktion von Farb- und Texturmerkmalen
- Erkennung einfacher Objekte und Szenen

Die Ergebnisse derartiger Analysen werden zur Kategorisierung der multimedialen Inhalte herangezogen. Die Ansätze zur Generierung von technischen und beschreibenden Metadaten umfassen die Erfassung und Harmonisierung der Inhalte, die Extraktion semantischer Information bis hin zu Methoden des maschinellen Lernens. Weil die Qualität von Suchergebnissen bei Rundfunkarchiven im Vordergrund steht, ist bei automatischen Indizierungsverfahren an die Fehlerfreiheit der Ergebnisse ein besonders hoher Anspruch zu stellen. Während Texterkennungssysteme (OCR) bereits heute die Dokumentare in Rundfunkanstalten unterstützen sind Sprach- und Sprechererkennung bislang allenfalls experimentell im Einsatz

Bei der Rückwärtsdigitalisierung im Rundfunk ist regelmäßig der Personaleinsatz der begrenzende Faktor für das umsetzbare Volumen. Eine Auslagerung der Metadatengenerierung an externe Dienst-

leister eröffnet die Aussicht, die oftmals wertvollen Medienbestände schneller einer Gewinn bringenden Zweitverwertung zuführen zu können. Die flexible Wahl von Dienstleistern versetzt das Archiv in die Lage, flexibel die qualitativ hochwertigere inhaltliche Erschließung in Anspruch zu nehmen.

Die in den Rundfunkanstalten praktizierte inhaltliche Erschließung ist subjektiv geprägt in ihrer Form, ihrem Umfang und ihrer Wortwahl. Die Ausrichtung und Zielsetzung der Rundfunkanstalt beeinflusst maßgeblich, inwieweit Sachinhalte, Bildinhalte und Rechte bei der inhaltlichen Erschließung eingearbeitet werden. Die Einbindung externer Dienstleister für die Analyse von Audio-, Sprach-, Bild- und Videodaten eröffnet die Möglichkeit, suchbare Metadaten und semantische Beschreibungen für die spezifischen Bedürfnisse von Fachdisziplinen und Nutzergruppen zu generieren.

Ingest eines Langzeitarchivs kann potenziell bei folgenden, rechenintensiven Funktionen von einer Auslagerung auf Rechenressourcen des Grids profitieren:

- Validierung von Dokumentformaten [Sch08]
- Extraktion und Generierung von technischen als auch beschreibenden Metadaten [Sch08]

Für *Access* eines Langzeitarchivs eröffnen sich für folgende Funktionen Potenziale für eine Auslagerung auf Rechenressourcen des Grid:

- Metadatenbasierte Suche [Sch08]
- Wissensbasierte Suche [Sch08]
- Generierung von Vorschauqualität zur Beschreibung
- Formatumwandlung angeforderter Archivobjekte bei Auslieferung

Data Management, Preservation Planning und Administration

Derzeit sind keine Potenziale für die Nutzung von *Grid-Resourcen* für die Archivaufgaben *Data Management, Preservation Planning* und *Administration* absehbar. Es ist aber vorstellbar und zu empfehlen, *Grid-Resourcen* dahingehend zu nutzen, Sicherheitsmechanismen effizient in jede dieser Archivaufgaben zu integrieren. Dies gilt auch für *Ingest* und *Access*.

7 Fazit

Das Ziel der vorliegenden Expertise mit dem Titel „*Vertrauenswürdige und abgesicherte Langzeitarchivierung multimedialer Inhalte*“ war zu beleuchten, welche Bedrohungen in digitalen Langzeitarchiven multimedialer Inhalte vorherrschen und warum die Sicherung der Verfügbarkeit, Integrität, Authentizität, Vertraulichkeit und Nachweisbarkeit digitaler Information notwendig ist. Im Mittelpunkt stand dabei, wie der Baustein IT-Sicherheit in bestehende und zukünftige Strategien und Konzepte der digitalen Langzeitarchivierung integriert werden kann.

Bei der digitalen Langzeitarchivierung multimedialer Inhalte steht man im Allgemeinen Gefährdungen gegenüber verbunden mit einem Informationsverlust. Gefährdungen sind in der Regel durch unvorhergesehene Geschehnisse wie technische Fehler oder Ausfälle oder gar eine Naturkatastrophe bestimmt. Solche Gefährdungen hat die Expertise nicht außer Acht gelassen, Schwerpunkt waren jedoch die zunehmenden Bedrohungen, die von menschlichen Handlungen mit gezielten Absichten ausgehen. Bedrohungen verursacht durch Manipulationen, wofür digitale Systeme und Multimedia eine breite Angriffsfläche bieten, werden in Zukunft verstärkt auf digitale Langzeitarchive zukommen. Daher ist es unabdingbar, IT-Sicherheit und Sicherheitsmechanismen in Planungen und Konzepte einzubinden.

Ausgehend von der Erhebung allgemeiner bestehender Anforderungen wurden in der Expertise anhand zweier Anwendungsszenarien (Hochschul-Medienzentren und Rundfunkanstalten) exemplarische Soll-Anforderungen für eine vertrauenswürdige und abgesicherte digitale Langzeitarchivierung aufgestellt. Auf Grundlage dieser Soll-Anforderungen ist der Ist-Zustand für die Anwendungsszenarien erhoben worden, um den Handlungsbedarf und die Möglichkeiten für die Integration von Sicherheitsmechanismen aufzuzeigen. Ziel war es darzulegen, wie vertrauenswürdige und abgesicherte Langzeitarchivierungssysteme konzipiert werden können, indem man IT-Sicherheit mit einbezieht.

In der Expertise wurde zunächst eine Auswahl existierender Studien mit Bedeutung für die vertrauenswürdige und abgesicherte Langzeitarchivierung aufgearbeitet und allgemeine Anforderungen und Annahmen aufgestellt. Die exemplarischen Langzeitarchivierungssysteme Hochschul-Medienzentren und öffentlich-rechtliche Rundfunkanstalten wurden allgemein charakterisiert und analysiert, wobei grundsätzliche Komponenten ihrer allgemeinen technischen Infrastruktur in Funktion und Struktur sowie die Informationsflüsse erhoben und beschrieben wurden. Dazu zählten Aspekte wie die technische Infrastruktur, Menge, Art und Ort anfallender Daten, eingesetzte technische Systeme, Art und Umfang der technischen Aufbereitung der Daten sowie Inhaltsbeschreibungen, Rechte-Daten und Metadaten. Darauf aufbauend erfolgte eine Systemabstraktion mit Zuordnung der Anforderungen und der Annahmen für die exemplarischen Szenarien als vertrauenswürdige und abgesicherte Langzeitarchivierungssysteme gemäß dem Referenzmodell. Im weiteren Verlauf der Expertise wurden Herangehensweisen zur Integration von Sicherheitstechnologien für eine vertrauenswürdige und abgesicherte Langzeitarchivierung multimedialer Inhalte evaluiert. Dazu wurden die Sicherheitsaspekte Verfügbarkeit, Integrität, Authentizität, Vertraulichkeit und Nachweisbarkeit allgemein dargestellt und die Auswirkungen von Verletzungen der Sicherheitsaspekte wurden beschrieben. Bedrohungen wurden aufgezeigt, Angriffe dargelegt, Sicherheitsrichtlinien und IT-Sicherheitsmanagement wurden erläutert. Weiterhin wurden Sicherheitsmechanismen eingeführt und ihre Aufgaben beschrieben. Im Zusammenhang mit der Integration von Sicherheitstechnologien für eine vertrauenswürdige und abgesicherte Langzeitarchivierung digitaler Information wurden die *Common Criteria* näher beleuchtet, Vertrauen und Vertrauenswürdigkeit wurden abgegrenzt und Dokumentation und Transparenz wurden gegenübergestellt.

Im weiteren Verlauf wurden die fünf Sicherheitsaspekte Verfügbarkeit, Integrität, Authentizität, Vertraulichkeit und Nachweisbarkeit mit ihrer Bedeutung im Kontext der digitalen Langzeitarchivierung multimedialer Inhalte beschrieben und Einsatzmöglichkeiten von Sicherheitstechnologien und deren Eignung zur Schaffung einer vertrauenswürdigen und abgesicherten Langzeitarchivierung wurde evaluiert. Anhand der fünf Sicherheitsaspekte wurden Soll-Anforderungen für eine vertrauenswürdige und abgesicherte Langzeitarchivierung multimedialer Inhalte aufgestellt, denen wiederum Sicherheitsmechanismen zugeordnet wurden. Basierend auf den existierenden Studien in Verbindung mit der erfolgten Erhebung der exemplarischen Langzeitarchivierungssysteme wurden Ist-Zustand und Soll-Anforderungen einander gegenübergestellt. So konnten letztendlich Aussagen über den Handlungsbedarf getroffen werden und die Sicherheitstechnologien konnten in Bezug auf die organisatorischen und technischen Rahmenbedingungen innerhalb der betrachteten Szenarien reflektiert werden. Letzteres erfolgte anhand einer Validierung an praktischen Beispielen.

Die Vorgehensweise der Expertise war Komponenten-orientiert, die Untersuchung der Integration von Sicherheit erfolgte demgemäß auf Grundlage

a. der Systemkomponenten, ausgehend von der Erhebung der Systemarchitektur, sowie

b. des digitalen Archivobjektes, ausgehend von der Erhebung der Informationsflüsse.

Die Sicherheitsanalyse geschah demnach sowohl Systemkomponenten-orientiert als auch Objekt-orientiert und zeigte dementsprechend den Handlungsbedarf gemäß diesem Ansatz auf. Eine szenarienabhängige und systemspezifische Analyse hätte demgegenüber eine Netzwerksicherheitsanalyse mit expliziten und detaillierten *Security Scans* erfordert. Dies war in dieser Expertise nicht vorgesehen. Ausgehend von der Erhebung der Systemarchitektur und der Informationsflüsse der Beispielszenarien Hochschul-Medienzentren und Rundfunkanstalten wurden mittels einer Abstraktion die anwendbaren Sicherheitsmechanismen über die Anwendungsszenarien hinweg bestimmt. Dadurch konnten allgemeingültige Aussagen über einen generellen Handlungs- und Standardisierungsbedarf getroffen werden. Grundsätzliches Ziel dieser Expertise war es, eine Grundlage für Erweiterung des bestehenden *nestor-Kriterienkatalogs* [Nes06], welcher allgemeingültig ist für alle Arten von Langzeitarchiven, im Punkt Sicherheit zu schaffen und Vorschläge für zukünftige Weiterentwicklungen bereitzustellen.

In dieser Expertise wurde festgestellt, dass die Anforderungen an eine vertrauenswürdige und abgesicherte Langzeitarchivierung digitaler Inhalte abhängig von Design, Konfiguration und Implementierung jedes einzelnen Langzeitarchivsystems sind. Dementsprechend ist auch der umgesetzte und vorhandene Ist-Zustand verschieden. Anforderungen für eine vertrauenswürdige und abgesicherte Langzeitarchivierung sollten in jedem Einzelfall für jedes Archiv entsprechend seiner Aufgaben und des enthaltenen Bestands in einer *Security Policy* festgehalten sein.

An dieser Stelle sei nochmals die Aufgabe bzw. Mission der Langzeitarchivierung und die Problematik der digitalen Langzeitarchivierung erwähnt:

Mit der *Langzeitarchivierung* soll Information von bleibendem wissenschaftlichem, künstlerischem oder gesellschaftlichem Wert [Bor03] erhalten bleiben, wobei der Zugriff und die Verfügbarkeit für eine bestimmte autorisierte Zielgruppe sowohl in der Gegenwart als auch in der Zukunft gewährleistet sein soll. Also über einen Zeitraum hinaus, innerhalb dessen technologische und soziokulturelle Veränderungen eintreten werden. In Bezug auf den Aspekt der Langzeit bedeutet dies die verantwortliche Entwicklung von Strategien, die den beständigen, vom Informationsmarkt verursachten Wandel bewältigen können [ScLi04]. Eine Information soll also auch in 100 Jahren noch verfügbar und lesbar sein. Laut *OAIS-Referenzmodell* [CCSDS02], auf welches sich diese Expertise grundlegend bezogen hat, bezeichnet ein Archiv eine Organisation zusammengesetzt aus Personen und Systemen, deren Aufgabe bzw. Verantwortung es ist, die Information zu erhalten und sie für eine bestimmte Zielgruppe zugänglich zu machen.

Die *digitale Langzeitarchivierung* bezieht sich auf Methoden und Strategien zur Erhaltung digitaler Information. Digitale Information unterliegt einem vom Informationsmarkt verursachten sehr schnellen Wandel. Die Schnelllebigkeit der technischen Weiterentwicklungen begründet dass die digitale Langzeitarchivierung vor neuen Herausforderungen und Anforderungen steht. So wird immer ein Abspielsystem [Bor03] benötigt, welches die in Zeichenströmen gespeicherte Information interpretiert und darstellt, um Zugang zur Information zu haben. Es gilt daher nicht nur das digitale

Archivobjekt selbst zu schützen, es muss ebenso sichergestellt werden, dass alle benötigten Systemressourcen verfügbar sind, so dass die Formate immer interpretierbar sind und der Zugang über ein benötigtes Abspielsystem gewährleistet ist. Dazu werden verschiedene Strategien verfolgt wie z.B. Gewährleisten der Auf- und Abwärtskompatibilität alter und neuer Systemumgebungen mittels Emulationen oder das Migrieren veralteter Formate, in denen die Information gespeichert ist, in neuere, in einer Systemumgebung abspielbare bzw. interpretierbare Formate.

Heutige digitale Archive bauen ihre technische Infrastruktur zunehmend aus angepasster Standard-Informationstechnik anstatt spezieller teuerer Gerätetechnik auf. Dadurch werden sie grundsätzlich anfällig für Schädlinge der Informationstechnik. Hinzu kommen Bedrohungen, die mit den Medien und Formaten für audiovisuelles Archivmaterial einhergehen, da diese vermehrt als Träger von Schadprogrammen identifiziert werden. Digitale Archive sind noch jung und eine Reihe von Problemstellungen wird sich mit fortdauerndem Betrieb neu aufzeigen. Wegen des geringen Alters der digitalen Archive war beispielsweise eine Migration der Archivbestände bislang nicht zwingend erforderlich. Mit der bei einem Formatwechsel einhergehenden Transkodierung werden Qualitätseinbußen befürchtet.

Verschlüsselung und digitale Signaturen werden in Bezug auf Langzeitarchivierung gerne als kritisch betrachtet. Es wird gefürchtet, dass mit Verwendung dieser Sicherheitsmechanismen die verschlüsselte Information auf lange Sicht gesehen nicht interpretiert werden kann und nicht verfügbar bzw. unzugänglich ist. Dies nämlich dann, wenn der Schlüssel verloren geht und das Verschlüsselte unlesbar bleibt, was unter dem Aspekt der Langzeit als wahrscheinlich bewertet wird. Hinzu kommt, dass eine derartige Beweissicherheit bisher nicht notwendig erschien in den in dieser Expertise betrachteten Langzeitarchiven. In medizinischen digitalen Langzeitarchiven oder im Bereich des E-Governments werden Verschlüsselungen auf Grundlage digitaler Signaturen bereits erfolgreich eingesetzt und sind zwingend erforderlich.

Grundsätzlich gestaltet sich die Einplanung von Sicherheit in digitale Langzeitarchive während ihres Neuaufbaus einfacher als durch nachträgliche Ergänzung in bestehende Systeme. Institutionen, welche sich traditionell mit der Bewahrung von Medieninhalten auseinandersetzen (wie hier Rundfunkanstalten und Hochschul-Medienzentren) haben bislang zwei wesentliche Entwicklungen durchlaufen: Erstens die Digitalisierung der Medienobjekte und zweitens die Digitalisierung und Vernetzung der Produktions- und Archivumgebung durch Umstellung auf Informationstechnik. Dabei wurden sowohl Medienobjekte, Infrastruktur als auch Verarbeitungsprozesse umgestellt. Der gesamte Wandel verläuft schrittweise und schreitet weiter voran. Der erstmalige Aufbau der digitalen Archive kann allgemein als bewältigt betrachtet werden. In dessen Zuge wurden, gemäß den damals bestehenden Anforderungen, ebenfalls Sicherheitsmechanismen implementiert. Exemplarisch seien hier *Uniform Resource Name (URN)* und *persistente Identifikatoren (PI)* benannt. Mit dem laufenden Einsatz der Archive und aufgrund zunehmender Bedrohungen haben sich die Anforderungen erweitert bzw. neu gestaltet. Digitale Langzeitarchive müssen Sicherheitsmechanismen zum heutigen Zeitpunkt und auch in Zukunft mehr denn je einbeziehen. Zukünftige zu erwartende Aus- und Umbaustufen bieten neue Gelegenheiten, Sicherheitsmechanismen einzuplanen.

Um digitale Langzeitarchive vertrauenswürdig und abgesichert zu gestalten, ist es in Zukunft unumgänglich, Sicherheitsmechanismen einzubeziehen. Sicherheitsmechanismen wie beispielsweise Verschlüsselungen, digitale Wasserzeichen oder Antivirenprogramme, die a priori eingesetzt werden, um Medienbrüchen vorzubeugen. Sicherheitsmechanismen können auch beispielsweise Hashfunktionen oder Verfahren der digitalen Medien- und Computerforensik zur Überprüfung auf Informationsveränderungen sein. Und es gibt auch Sicherheitsmechanismen, wie beispielsweise Biometrie oder digitale Signaturen zur Authentifizierung. Nur so kann zukünftigem Schaden vorgebeugt, Informationsverlust festgestellt, eine Manipulation nachgewiesen und Originalinformation möglicherweise rekonstruiert werden.

Für die tatsächliche Umsetzung dieser Forderung in der Praxis sind Lösungen gesucht, welche eine effiziente Integration von Sicherheitsmechanismen in zukünftige digitale Langzeitarchive ermöglichen, ohne große Aufwände und Kosten zu verursachen. Hier sind neue Paradigmen für die Organisation von Informationsverarbeitung wie beispielsweise Service-orientierte Architektur (SOA) oder Grid-Computing richtungweisend. Sie bieten zukünftig die Möglichkeit bestimmte Sicherheitsmechanismen wie beispielsweise digitale Signaturen als Dienste einzubinden. Jedoch darf darüber

nicht die Sicherheit der eigentlichen Infrastruktur der digitalen Langzeitarchivierungssysteme vernachlässigt werden.

Zukünftige vertrauenswürdige und abgesicherte digitale Langzeitarchivierungssysteme haben folgende *Mindestanforderungen* zu erfüllen, wie in Kapitel 5 eingehend beschrieben wurde.

- Verfügbarkeit
  - Archivobjekte
  - Systemressourcen (Hard- und Software sowie Netzwerk)
  - Darstellungsanwendung
- Integrität
  - Archivobjekte
  - Systemressourcen (Hard- und Software sowie Netzwerk)
  - Darstellungsanwendung
- Authentizität/ Authentifizierungen
- Vertraulichkeit
- Nachweisbarkeit
- Wiederauffindbarkeit der Archivobjekte
- Interoperabilität
- Modularisierung und Erweiterbarkeit
- Plattformunabhängigkeit/ Flexibilität
- Transparenz
- Nachhaltigkeit

Die Gewährleistung der langfristigen Verfügbarkeit verwahrter digitaler Archivobjekte gestaltet sich komplex, ist für sich allein genommen nicht ausreichend und mit den anderen Sicherheitsaspekten eng verbunden. So ist die Kernaufgabe digitaler Langzeitarchivierungssysteme die Sicherstellung von Integrität und Authentizität der anvertrauten Archivobjekte, aber auch der Ressourcen und Darstellungsanwendungen. Ein digitales Langzeitarchiv muss abgesichert sein gegenüber Manipulationen und Angriffen. Aktuelle Entwicklungen wie die zunehmende Vernetzung der Systeme durch das Internet oder die wachsende Vielfalt von Medientypen und deren Verbindung zu Multimedia erschweren diese Aufgabe. Ein digitales Archivobjekt unterliegt unzähligen Medienwechseln, die oftmals mit Informationsverlust verbunden sind. Die Bestandserhaltung gestaltet sich hier umfangreicher und anders als in herkömmlichen Archiven. Die digitale Langzeitarchivierung ist ein dynamischer Prozess, der die ständigen Veränderungen und technischen Weiterentwicklungen berücksichtigen und dabei gleichzeitig den Bestand sichern muss. Dies ist nur zu realisieren durch angemessene Erhaltungsstrategien mit einer dynamischen Einbindung gezielter Sicherheitsmaßnahmen.

Digitale Langzeitarchive sollten vor allem nachhaltig entwickelt werden. *Nachhaltigkeit* im Kontext der digitalen Langzeitarchivierung bedeutet, dass zum einen Strategien angewandt werden, um den bestehenden Archivbestand sicher zu erhalten, während zum anderen gleichzeitig Aufwand betrieben wird, den Bestand sinnvoll und angemessen zu erweitern. Dies schließt die Berücksichtigung und Erfüllung aller zuvor genannten Mindestanforderungen ein. Die Lösung liegt einerseits in der Anwendung von bestehenden Standards und Sicherheitsmechanismen sowie andererseits in der gleichzeitigen Weiterentwicklung eben dieser, denn heute entwickelte und eingesetzte Systeme sollen in Zukunft erweiterbar und an sich ändernde Gegebenheiten, unter anderem technische, rechtliche und kulturelle, anpassbar sein.

So unterstützt beispielsweise der Einsatz von Open Source und standardisierten Schnittstellen die einfache Erweiterung eines digitalen Langzeitarchivs mit der An- und Einbindung von weiteren Archivbeständen. Grid-Technologien unterstützen darüber hinaus die effiziente Einbindung von Sicherheitsmechanismen. So kann neben der Sicherheit und Vertrauenswürdigkeit auch die Plattformunabhängigkeit und Flexibilität gewährleistet werden. Dies ist in Zukunft von besonderer Relevanz, da

neben der Interaktion mit externen Akteuren, wie Produzent oder Konsument, vor allem die Interaktion mit Dienstleistern für (Web-)Services sowie der archivübergreifende Austausch stark zunehmen wird. Letztendlich können so die Vorteile des dynamischen Prozesses ständiger technologischer Weiterentwicklungen für die digitale Langzeitarchivierung genutzt werden. Ihre Aufgabe, den Austausch von Information zu ermöglichen und Information für eine lange Zeit unabhängig von zur Verfügung stehender Hard- und Software zugänglich zu machen, können digitale Langzeitarchive damit um so besser erfüllen.

Kernaufgaben digitaler Langzeitarchive sind und bleiben die langfristige Bewahrung des Bestands und seine Erweiterung. Bedingt durch die Auswirkungen moderner Informations- und Kommunikationstechnologie sind die Produktion, die Verbreitung, die Nutzung und die Bewahrung multimedialer Inhalte und Medien bereits seit geraumer Zeit tief greifenden Veränderungen unterworfen. Schon heute ist diese Entwicklung gekennzeichnet durch eine Beschleunigung der Produktions- und Wiederverwendungszyklen sowie eine rasant zunehmende Fülle und Vielfalt der bereitstehenden und umgesetzten Medieninhalte. Verstärkt durch den allgemeinen Trend zur Globalisierung sowie aus Gründen der Wirtschaftlichkeit wird es in Zukunft vermehrt zu organisationsübergreifenden Arbeitskonzepten und –Durchführungen kommen. Dies bedeutet Umstrukturierungen der organisatorischen Verantwortlichkeiten und ist mit dem Ziel verbunden, verteilt bestehende Ressourcen durch effiziente Verknüpfung nutzen zu können. So wird es neue Arbeitsabläufe geben, wie die Zusammenarbeit über Organisationsgrenzen und technische Plattformen hinweg. Digitale Langzeitarchive sind Teil dieser Entwicklung mit der Folge, dass sie enger an die Produktions- und Wiederverwendungsprozesse gekoppelt sind. Mit solch einer zukünftigen Öffnung der Archive mit all ihren Konsequenzen wie einer Vielzahl von verschiedenen Nutzern, einer Vielzahl von technischen Systemen, einem erhöhten Zugriff, einer erhöhten Datendurchsatzrate, einem erhöhten Volumen, einer Vielzahl von Verwendungszwecken, usw., sehen sich die Archive einem neuen Umfeld und neuen Herausforderungen ausgesetzt. Einerseits werden Sicherheitsmechanismen benötigt, um die Kernaufgaben des Archivs zu schützen. Andererseits sind Sicherheitsmaßnahmen essenziell, welche sich mit dem neuen Umfeld auseinandersetzen.

Bestimmendes Mittel für eine Gewährleistung der Sicherheit in den Archiven heute ist der Schutz der Infrastruktur und des gesicherten, authentifizierten Zugangs zu dieser, wozu umfassende organisatorische und technische Maßnahmen erarbeitet wurden bzw. bereitstehen. Mit einer zunehmenden Auflösung der Grenzen des Archivs, wie beispielsweise durch die Nutzung externer Dienstleister für Funktionsentitäten des Archivs vermittels Grid-Technologien, rückt darüber hinaus die Sicherung des Archivobjekts in den Fokus, denn diese muss nun auch innerhalb nicht vertrauenswürdiger Umgebungen erhalten und gewährleistet werden. Hierzu wird eine Schwerpunktverlagerung weg von der Sicherung der Institution des Archivs und seiner Infrastruktur als eine vertrauenswürdige Umgebung hin zu einer Sicherung der einzelnen Archivobjekte für die nicht vertrauenswürdigen Umgebungen erforderlich.

Dieser Expertise lag das *OAIS-Referenzmodell* zugrunde, welches ein hervorragendes Rahmenwerk darstellt für den Aufnahmeprozess, Bereitstellungsprozess, Erhaltungsprozess, die Architektur und Infrastruktur, Informationsflüsse, das digitale Archivobjekt sowie Repräsentation und Handling desselbigen, Metadaten, Rechtemanagement und Erhaltungspolicies. Es ist also ein Rahmenwerk für ein abgeschlossenes digitales Archiv. In zukünftigen Szenarien wird es jedoch zunehmend zu einem archivübergreifenden Informationsaustausch, einer Verteilung der Ressourcen sowie zu einer vermehrten externen Kommunikation und Interaktion kommen. Solche Prozesse, resultierend aus der Interaktion mit externen Parteien, wie beispielsweise Dienstleister oder anderen Archiven, werden mit Bezug auf die Inhaltsobjekte verarbeitenden Einheiten in dem derzeitigen OAIS-Referenzmodell nicht ausdrücklich berücksichtigt.

Weiterhin bezog sich die Expertise auf die Beschreibung des digitalen Objekts von Thibodeau [Thib02], welcher den Einbezug der Semantik in seinen Ausführungen vernachlässigt. Die Semantik spielt im Zusammenhang mit der Sicherung der Integrität und Authentizität digitaler Information jedoch eine wichtige Rolle, was gerade in digitalen Langzeitarchiven von entscheidender Bedeutung ist. So können die Bedeutung, Zusammenhänge und Beziehungen von archivierter Information zum Zeitpunkt der Archivierung und zum Zeitpunkt der Bereitstellung sehr unterschiedlich sein. Gesetzliche, technische oder kulturelle Gegebenheiten ändern sich auf lange Sicht gesehen. Information muss zukünftig verstärkt in ihrer Semantik, im Kontext ihrer Anwendung, erfasst sein, um

analysiert werden zu können. Dies spielt eine zentrale Rolle bei der Einbindung von Sicherheitsmechanismen in digitalen Langzeitarchivierungssystemen, was ohne die Berücksichtigung der Semantik nicht ausreichend erfolgen kann.

Handlungsbedarf besteht vor allem auch im Bereich der Grid-Technologien und der verteilten Speicherlösungen, da diese in zukünftigen Systemen zur digitalen Langzeitarchivierung vermehrt zum Einsatz kommen werden. Darüber hinaus müssen die Sicherheitsmechanismen selbst auch weiterentwickelt werden.

Als Fazit lassen sich vier Kernaussagen in Bezug auf eine vertrauenswürdige und abgesicherte digitale Langzeitarchivierung multimedialer Inhalte formulieren:

1. Sicherheitsmaßnahmen sind zukünftig in Systemen zur digitalen Langzeitarchivierung multimedialer Inhalte unabdingbar, da es neben unbeabsichtigten Vorkommnissen und organisatorischen Mängeln verstärkt Bedrohungen aufgrund von Manipulationen bzw. gezielten Angriffen geben wird.

2. Die Sicherung der Verfügbarkeit, Integrität, Authentizität, Vertraulichkeit und Nachweisbarkeit muss für ein gegenwärtiges digitales Langzeitarchiv einschließlich seines Bestands gewährleistet sein, aber auch der Nachweis dieser Eigenschaften zu einem späteren Zeitpunkt muss sichergestellt werden.

3. Als Trend im Bereich der digitalen Langzeitarchive steht eine Entwicklung weg von einer geschlossenen Archivumgebung einschließlich seiner Systeme hin zu deren Öffnung mit der Nutzung verteilt vorliegender Ressourcen und der Einbindung externer Dienste zu erwarten.

4. Aufgrund der Öffnung der Archive gewinnt in Zukunft die Sicherung des einzelnen Archivobjekts neben der Sicherung der Infrastruktur des Archivs an Bedeutung, da ein Archivobjekt zunehmend auch in unbekannten Umgebungen gesichert sein muss. Hier spielt neben der Sicherung der Integrität und Authentizität die Vertrauenswürdigkeit eine besondere Rolle.

Generell ist festzuhalten, dass derzeit eine Vielzahl von Aktivitäten in Bezug auf Standardisierungen läuft, die eine *vertrauenswürdige und abgesicherte digitale Langzeitarchivierung* betreffen und dort bereits angewendet werden bzw. potentiell angewendet werden können. So werden mit den Entwicklungen in den Bereichen SOA und Grid-Technologien Ansätze geboten, Sicherheitsmechanismen möglichst einfach und effizient in die digitale Langzeitarchivierung zu integrieren. Der Handlungsbedarf besteht nun darin, die Verwendbarkeit der neuen Techniken (Grid, SOA) dahingehend zu überprüfen, inwiefern sie in bestehende Systeme und Infrastrukturen eingebunden werden können. Hier muss eine Abwägung der Synergiepotentiale erfolgen, um Auslagerungsmöglichkeiten zu evaluieren und bedienen zu können.

Zukünftige Aktivitäten sollten vermehrt im Sinne der Nachhaltigkeit stattfinden und Mechanismen der Sicherheit standardisiert in die Abläufe eines digitalen Langzeitarchivs einbeziehen, um eine vertrauenswürdige und abgesicherte Langzeitarchivierung zu realisieren. Denn die Herausforderung der digitalen Langzeitarchivierung ist der Aspekt der *Langzeit*, insbesondere die Frage, wie sichergestellt werden kann, dass Information von bleibendem wissenschaftlichem, künstlerischem oder gesellschaftlichem Wert [Bor03] erhalten werden kann, wobei der Zugriff und die Verfügbarkeit für eine bestimmte autorisierte Zielgruppe sowohl in der Gegenwart als auch in der Zukunft gewährleistet ist. Also über einen Zeitraum hinaus, innerhalb dessen technologische und soziokulturelle Veränderungen eintreten werden. Nicht anders als durch die Sicherung der Verfügbarkeit, Integrität, Authentizität, Vertraulichkeit und Nachweisbarkeit wird man der Anforderung und Verantwortung gerecht, nachhaltige Strategien zu entwickeln, die den beständigen, vom Informationsmarkt verursachten Wandel bewältigen können, um multimediale Inhalte von bleibendem Wert zu erhalten.

Literaturverzeichnis

[Alt05] R. Altenhöner: Daten für die Zukunft – Das BMBF-Projekt Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen (kopal) und seine Hintergründe. Bibliothek 29, Nr. 2, 2005.

[ARD07] ARD. ARD beschließt Strategie für die digitale Medienwelt - Tagesschau aufs Handy. Pressemeldung 19.06.2007. URL: http://www.ard.de/-/id=620988/sbkf0v/index.html

[Bar02] S. Barman: Writing Information Security Policies. New Riders Publishing, First Edition, 2002.

[BR04] Bayerischer Rundfunk: Das Fernseharchiv. 28.09.2004. URL: http://www.br-online.de/br-intern/thema/tag-der-archive-2006/fernseharchiv.xml

[BaAr05] A. Bazaz und J. D. Arthur: On Vulnerabilities, Constrains and Assumptions. In: ArXiv Computer Science e-prints, 2005.

[Bis03] M. Bishop: Computer Security – Art and Science. Boston, Addison-Wesley, 2003.

[Bor03] U.M. Borghoff. Standards und Standardisierung im Kontext von Grid-Technologien und Langzeitarchivierung. Studie der AG Grid des Kompetenznetzwerks nestor, in Vorbereitung. Universität der Bundeswehr München, München, 2003.

[Bor05] U.M. Borghoff. nestor-materialien 3 - Vergleich bestehender Archivierungssysteme, Hrsg. Von Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland, Frankfurt am Main: nestor c/o Die Deutsche Bibliothek, 2006. URL: http://nbn-resolving.de/urn:nbn:de:0008-20050117016

[Bor08] U.M. Borghoff. Standards und Standardisierung im Kontext von Grid-Technologien und Langzeitarchivierung. Studie der AG Grid des Kompetenznetzwerks nestor, in Vorbereitung. Universität der Bundeswehr München, München, 2008.

[BRSS03] U. M. Borghoff, P. Röding, J. Scheffczyk, L. Schmitz: Langzeitarchivierung; Methoden zur Erhaltung digitaler Dokumente. dpunkt.verlag Heidelberg, 2003.

[BSI06] Bundesamt für Sicherheit in der Informationstechnik (BSI): IT-Grundschutz-Kataloge. Stand 2006. URL: http://www.bsi.bund.de/gshb/index.htm

[BSI08] Bundesamt für Sicherheit in der Informationstechnik (BSI): BSI-Standards. 2008. URL: http://www.bsi.de/literat/bsi_standard/index.htm

[BVG07] BVerfG, 1 BvR 2270/05 vom 11.9.2007, Absatz-Nr. (1-213), 2007. URL: http://www.bverfg.de/entscheidungen/rs20070911_1bvr227005.html

[Chr06] H.-P. Christmann. Intelligente Speicherlösungen für Broadcast-Anwendungen. In [FKT] 2006/10, S. 608-613, 2006.

[CCSDS02] Consultative Committee for Space Data Systems (CCSDS): Reference Model for an Open Archival Information System (OAIS). Recommendation for Space Data System Standards, CCSDS 650.0-B-1, BLUE BOOK, January 2002. URL: http://public.ccsds.org/publications/archive/650x0b1.pdf

[CC05] Common Criteria CC: Common Criteria for Information Technology Security Evaluation; Part 1: Introduction and general model. August 2005, Version 2.3, CCMB-2005-08-001, 2005.

[CPB07] Corporation for Public Broadcasting (CPB). Public Broadcasting Metadata Dictionary Project (PBCore). Homepage, URL: http://www.pbcore.org/

[Coy06] W. Coy. nestor-materialien 5 - Perspektiven der Langzeitarchivierung multimedialer Objekte, Hrsg. Von Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland, Frankfurt am Main: nestor c/o Die Deutsche Bibliothek, 2006. URL: http://nbn-resolving.de/urn:nbn:de:0008-20051214015

[DDNC07] The Digital Curation Center, DigitalPreservationEurope, nestor, Center for Research Libraries: Core Requirements for Digital Archives. 2007. URL: http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=92

[Dit00] J. Dittmann. Digitale Wasserzeichen: Grundlagen, Verfahren, Anwendungsgebiete. Springer-Verlag Berlin Heidelberg New York, 2000.

[Dit04] J. Dittmann: IT-Security. 2004.

[DOM05] DOMEA Konzept URL: http://www.kbst.bund.de/cln_028/nn_838516/SharedDocs/Anlagen-kbst/Domea/domea-organisationskonzeptes-2-1,templateId=raw,property=publicationFile.pdf/domea-organisationskonzeptes-2-1.pdf

[Ebn05] Ebner, A.: Austausch von Metadaten – Broadcast Metadata exchange Format, BMF. Institut für Rundfunktechnik (IRT), Technischer Bericht Nr. B 193/2005, 2005. URL: http://www.irt.de/IRT/publikationen/BlaueBerichte/Blauer%20%20Bericht%20193%20Ebner%2026.9.2005.pdf

[EK05] Ebner, A. und Knör, R.: Fortschritte bei der digitalen filebasierten Produktion und in der Archivierung. In Jahresbericht 2005, Institut für Rundfunktechnik (IRT), S. 14-17, 2005.

[Eck03] C. Eckert. IT-Sicherheit Konzepte – Verfahren – Protokolle. 2. Auflage, Oldenburg Wissenschaftsverlag, 2003.

[Eck06] C. Eckert: IT-Sicherheit: Konzepte – Verfahren – Protokolle. Oldenbourg Wissenschaftsverlag München Wien, 2006.

[FKT] FKT - Fachzeitschrift für Fernsehen, Film und elektronische Medien sowie Verbandszeitschrift der Fernseh- und Kinotechnischen Gesellschaft (FKTG). Schiele & Schön, Berlin. ISSN 1430-9947

[Far08a] H. Farid: Digital Image Forensics. American Academy of Forensic Sciences, Washington, DC, 2008

[Far08b] H. Farid: Digital Video Forensics. American Academy of Forensic Sciences, Washington, DC, 2008

[FoKe04] I. Foster, C. Kesselman: The Grid: Blueprint for a New Computing Infrastructur., 2. Auflage, Elsevier, o. O., 2004.

[FrLG06] J. Fridrich, J. Lukas, M. Goljan: Digital Camera Identification from Sensor Noise. In: IEEE Transactions on Information Security and Forensics, June 2006, Vol. 1(2), S. 205-214, 2006.

[HO07] Heise Online. ARD will nicht nur programmbegleitend im Internet aktiv sein. Meldung 11.10.2007, URL http://www.heise.de/newsticker/meldung/97268

[HoLo03] John D. Howard, Thomas A. Longstaff: A Common Language for Computer Security Incidents (SAND98-8667); Sandia National Laboratories. 1998 (0-201-63346-9), Forschungsbericht, 1998.

[JoFa07] M.K. Johnson and H. Farid: Exposing Digital Forgeries in Complex Lighting Environments. In: IEEE Transactions on Information Forensics and Security, 2(3), S. 450-461, 2007.

[KaRo97] U. Kampffmeyer, J. Rogalla: Grundsätze der elektronischen Archivierung. VOI-Kompendium Band 3. VOI Verband Organisations- und Informationssysteme e. V., Darmstadt 1997, ISBN 3-932898-03-6, 1997.

[KLD07] Stefan Kiltz, Andreas Lang, Jana Dittmann; Taxonomy for Computer Security Incidents; In: Cyber Warfare and Cyber Terrorism; Publisher: Information Science Reference (IGI Global); Editors: Lech J. Janczewski, Andrew M. Colarik; pp. 412-417; ISBN 978-1-59140-991-5; 2007

[KMM+06] N. Khanna, A. K. Mikkilineni, A. F. Martone, G. N. Ali, G. T.-C. Chiu, J. P. Allebach, E. J. Delp: A survey of forensic characterization methods for physical devices. In: Proceedings of the 6th Digital Forensics Research Workshop (DFRWS), Lafayette, Indiana, August 2006, S. 17-28, 2006.

[Klu08] J. Klump. Anforderungen von e-Science und Grid-Technologie an die Archivierung wissenschaftlicher Rohdaten. Studie der AG Grid des Kompetenznetzwerks nestor, in Vorbereitung. GeoForschungszentrum Potsdam, Potsdam, 2008.

[KK05] H. Krömker, P. Klimsa. Handbuch der Medienproduktion – Produktion von Film, Fernsehen, Hörfunk, Print, Internet, Mobilfunk und Musik. Verlag für Sozialwissenschaften, Wiesbaden, 2005.

[LDKH06] A. Lang, J. Dittmann, S. Kiltz, T. Hoppe: Future Perspectives: The Car and its IP-Address - A Potential Safety and Security Risk Assessment. In: Computer Safety, Reliability, and Security, Proceedings of the 26th International Conference SAFECOMP 2007, 18.-21. September 2007,

Nuremberg, Germany, Springer LNCS 4680, S. 40-53, Editors: Francesca Saglietti, Norbert Oster, 2007.

[LoC05] Library of Congress: METS: Überblick und Anleitung (Übersetzung: Angelika Menne-Haritz) Juli 2005. URL: http://www.loc.gov/standards/mets/METSOverview.v2_de.html

[LuFG05] J. Lukas, J. Fridrich, M. Goljan: Determining Digital Image Origin Using Sensor Imperfections. In: Proc. SPIE Electronic Imaging San Jose, CA, January 16-20, S. 249-260, 2005.

[LRF04] S. Lyu, D. Rockmore and H. Farid: A Digital Technique for Art Authentication. In: Proceedings of the National Academy of Sciences, 101(49), S. 17006-17010, 2004.

[MT04] A. Mauthe, P. Thomas. Professional Content Management Systems: Handling Digital Media Assets. Wiley, Chichester, 2004.

[Mei03] B. Meinschein: Intel / TCG, TCG-Symposium, Juli 2003.

[MHEG93] MHEG: Information Technology – Coded Representation of Multimedia and Hypermedia Information (MHEG), Part 1: Base Notation (ASN.1). Committee draft ISO/IEC CD 13522-1, June 1993. ISO/IEC JTC1/SC29/WG12, 1993.

[MAC+05] A.K. Mikkilineni, O. Arslan, P.-J. Chiang, R.M. Kumontoy, J.P. Allebach, G.T.-C. Chiu, et al.: Printer forensics using svm techniques. In: Proceedings of the IS&T's NIP21: International Conference on Digital Printing Technologies, Baltimore, MD, October 2005, Vol. 21, S. 223-226, 2005.

[MCA+05] A. K. Mikkilineni, P.-J. Chiang, G. N. Ali, G. T. C. Chiu, J. P. Allebach, Edward J. Delp III: Printer identification based on graylevel co-occurrence features for security and forensic applications. In: Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Con-tents VII, SSWMC, San Jose, California, USA, January 17-20, vol. 5681, S. 430-440, 2005.

[Mür05] S. Mürl. Redaktionsarbeit im privaten Fernsehen. In [KK05] S. 171-180, 2005.

[Nes04] Rechtsanwälte Goebel und Scheller (Bad Homburg v.d.H.). Expertise nestor-Digitale Langzeitarchivierung und Recht. Frankfurt am Main : nestor c/o Die Deutsche Bibliothek, 2004. URL: http://www.langzeitarchivierung.de/downloads/mat/nestor_mat_01.pdf

[Nes06] nestor-materialien 8 - Kriterienkatalog vertrauenswürdige digitale Langzeitarchive (dt./engl.), Hrsg. von der nestor-Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung, Frankfurt am Main: nestor c/o Die Deutsche Bibliothek, 2006. URL: http://edoc.hu-berlin.de/series/nestor-materialien/2006-8/PDF/8.pdf

[Nes06a] nestor-memorandum: Memorandum zur Langzeitverfügbarkeit digitaler Informationen in Deutschland, 2006. URL: http://www.langzeitarchivierung.de/downloads/memo2006.pdf

[Neu05] Neuroth, H. (nestor): Das Aufgabenfeld „Langzeitarchivierung" im bibliothekarischen Kontext – organisatorisch, technisch, juristisch. 94. Deutsche Bibliothekartag Düsseldorf, 2005. URL: http://www.langzeitarchivierung.de/downloads/2005_03_neuroth

[NN01] NN. Studie, Institut für Rundfunktechnik (IRT), 2001.

[OeDi05] A. Oermann, J. Dittmann: Evaluierung von Sicherheitsmechanismen für e-Learning zur Sicherheit der Integrität und Authentizität bei Medienübergängen. In: Marktplatz Internet: Von e-Learning bis e-Payment, Jantke, Klaus P. (Hrsg.); Fähnrich, Claus Peter (Hrsg.); Wittig, Wolfgang S. (Hrsg.) -Proceedings, Bonn: Ges. für Informatik, S. 391-399 (GI-Edition: Proceedings, 13. Leipziger Informatiktage, LIT 2005 Leipzig, 21.-23. September, 2005.

[OeDi06] A. Oermann, J. Dittmann: Trust in E-Technologies. In: Encyclopedia of E-Commerce, E-Government and Mobile Commerce, Mehdi Khosrow-Pour (Ed.) Information Resources Management Association, USA, Idea Group Reference, Hershey London Melbourne Singapore, S. 1101-1108, 2006.

[OLD05] A. Oermann, A. Lang und J. Dittmann: Verifier-Tuple for Audio-Forensics to Determine Speaker Environment. In: City University of New York (Veranst.): Multimedia and Security, MM&Sec'05 Proceedings, New York, NY, ACM, 2005, S. 57-62, Workshop New York, NY, USA August 1-2, 2005.

[Ott05] B. Otto. Format im Fernsehen. In [KK05], S. 165-170, 2005.

[Pfl03] C. Pfleeger und S.-L. Pfleeger: Security in Computing. Prentice Hall, 2003.

[PoFa04] A.C. Popescu and H. Farid: Statistical Tools for Digital Forensics. 6th International Workshop on Information Hiding, Toronto, Canada, 2004.

[PREMIS05] Data Dictionary for Preservation Metadata; Final Report of the PREMIS Working Group. Prepared on behalf of Preservation Metadata: Implementation Strategies (PREMIS), a working group jointly sponsored by OCLC and RLG, 2005. URL: http://www.oclc.org/research/projects/pmwg/premis-final.pdf

[RWF05] Regelwerk Fernsehproduktion, Metadaten für Austausch. Empfehlung der FSBL-K, Ausgabe März 2005.

[Röd07] J. Röder. Das Material Exchange Format (MXF) im netzwerkbasierten TV-Studio. In [FKT] 2007/01-02, S. 23-27, 2007.

[RStV06] Rundfunkstaatsvertrag (RStV), Staatsvertrag über den Rundfunk im vereinten Deutschland vom 31. August 1991, zuletzt geändert durch den Neunten Rundfunkänderungsstaatsvertrag vom 31. Juli bis 10. Oktober 2006.

[San05] K. Sandig. Fernsehtechnik Gestern und Heute. In [KK05], S. 110-126, 2005.

[Sau02] D. Sauter. Rechteauskunft. In Jahresbericht 2002, Institut für Rundfunktechnik (IRT), S. 22, 2002.

[Sau07] D. Sauter. Film und Fernsehen - Machtverschiebung durch Technik. In [FKT] 2007/06, S. 295-300, 2007.

[Sau07b] D. Sauter. Metadaten-Repository als Industriestandard, BMF - Broadcast Metadata exchange Format - Voraussetzungen für einen effektiven Metadatenaustausch. Präsentation auf der 2. Sitzung der nestor-Arbeitsgruppe Medien, Berlin, 22. November 2007.

[Sch05] R. Schäfer. Zukunftsperspektiven in der Fernsehtechnik. In [KK05], S. 127-138, 2005.

[Sch08] W. Schiffmann. Synergiepotenziale zwischen GRID- und e-Science-Technologien für die Langzeitarchivierung. Studie der AG Grid des Kompetenznetzwerks nestor, in Vorbereitung. FernUniversität in Hagen, Hagen, 2008.

[ScLi04] U. Schwens, H. Liegmann: Langzeitarchivierung digitaler Ressourcen. In: Grundlagen der praktischen Information und Dokumentation / begr. von Klaus Laisiepen, Ernst Lutterbeck u. Karl-Heinrich Meyer-Uhlenried. 5. völlig neu gefasste Ausg. - München : Saur. Bd. 1: Handbuch zur Einführung in die Informationswissenschaft und –praxis, S. 567- 570, 2004.

[StNa02] Steinmetz, R., Nahrstedt, C.: Multimedia Fundamentals Volume 1: Media Coding and Content Processing. IMSC Press Multimedia Series, Andrew Tescher (Series Ed.), Prentice Hall PTR, 2002.

[Str04] H. Strass. Grid Computing - Grid Storage. Gastbeitrag 10.12.2004, URL http://www.speicherguide.de/magazin/storagegrids.asp?theID=745

[Thi02] K. Thibodeau: Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. Council on Library and Information Resources: The State of Digital Preservation: An International Perspective, Conference Proceedings, July 2002. http://www.clir.org/pubs/reports/pub107/thibodeau.html

[WG06] B. Walsh, T. Geppert. Storage Management – digitale Inseln verbinden. In [FKT] 2006/12, S. 750-754, 2006.

[WaFa07] W. Wang and H. Farid Exposing Digital Forgeries in Interlaced and De-Interlaced Video. In: IEEE Transactions on Information Forensics and Security, 2(3), S. 438-449, 2007.

[WiSa03] H. Wilkens, D. Sauter: Informationstechnik in der Fernsehproduktion „IT-based Production“ und Havariekonzepte. In [FKT] 2003/8-9, S. 386-394, 2003.

[Wri07] R. Wright. Anforderungen an digitale Archive. In [FKT] 2007/03, S. 131-136, 2007.