



AG Archivierung der Kommission „Archäologie und Informationssysteme“  
im Verband der Landesarchäologen der Bundesrepublik Deutschland

# Ratgeber zur **Archivierung digitaler Daten**



**Trittsteine auf dem Weg zum Digitalarchiv**

Version 1.0.

Die AG Archivierung der Kommission „Archäologie und Informationssysteme“ im Verband der Landesarchäologen Deutschlands verfolgt das Ziel, allgemein verfügbare methodische und technische Ansätze zur Archivierung digitaler archäologischer Daten zu begutachten sowie praxisorientierte Empfehlungen, Vorschläge und Tipps zu geben. Die Ergebnisse dieser Arbeit münden in diesen Ratgeber. Aus verschiedenen Gründen sind einige Passagen dieses Ratgebers ergänzungswürdig geblieben. An diesen Stellen findet man jedoch einige stichpunktartige Hinweise zu den vorgesehenen Themen, als Anregung für weitere Recherchen oder einfach zum Nachdenken.

## Inhalt

1.	Einleitung.....	3
2.	Grundsätze.....	5
3.	Organisation der Archivierung.....	6
4.	Archivierungsrichtlinien.....	7
4.1.	Archivwürdigkeit.....	7
4.2.	Qualitätsmanagement.....	7
4.3.	Schutz der Informationen.....	7
5.	Metadaten.....	9
5.1.	Dublin-Core-Standard.....	9
5.2.	Metadaten für Archivierungsaufgaben.....	9
5.3.	Dokumentenmanagement.....	10
6.	Archivierungsmethoden.....	11
6.1.	Hardcopy.....	11
6.2.	Erhaltung.....	12
6.3.	Migration.....	12
6.4.	Emulation.....	13
6.5.	Sonderfall Datenbanken.....	13
6.6.	Notfall und Einzelfall.....	13
6.7.	Zusammenfassung der Methoden.....	14
7.	Datenträger und Datenformate.....	15
7.1.	Texte.....	15
7.2.	Bilder.....	16
7.3.	CAD-Daten.....	18
7.4.	Geodaten.....	19
7.5.	Datenbanken.....	23
8.	Erforderliche Ressourcen.....	25
9.	Archivierungserfahrungen.....	25
10.	OAIS - Open Archival Information System.....	26
10.1.	Datenübernahme (Ingest).....	26
10.2.	Datenaufbewahrung (Storage).....	26
10.3.	Datenverwaltung (Management).....	26
10.4.	Zugriff (Access).....	27
10.5.	Erhaltungsplanung (Preservation Planning).....	27
10.6.	Systemverwaltung (Administration).....	27
11.	Schlussfolgerungen, Empfehlungen.....	28
12.	Quellen/Verweise.....	29
Anlage 1:	Datei-Spezifikationen.....	30
Anlage 2:	Erstellung von Bildern für Bildarchive.....	32

**Sachstand April 2012,**  
(kleinere Ergänzungen: November 2015)

**Autoren:**  
David Bibby  
Andreas Brunn  
Sigmar Fitting  
Reiner Göldner  
Axel Posluschny

Entwurf:  
- 06.10.2010, R. Göldner  
Ergänzungen/ Korrekturen:  
- 13.11.2010, AG Archivierung  
- 17.02.2011, AG Archivierung  
- April/August 2011, R. Göldner  
- 24.08.2011, AG Archivierung  
- November 2011, R. Göldner  
- Februar 2012, AG Archivierung  
- 10.04.2012, R. Göldner  
Finalisierung:  
- 02.11.2015, R. Göldner  
Redaktion  
- R. Göldner



# 1. Einleitung

(Reiner Göldner)

„Digital“ ist Trend. Digitale Helfer findet man heute für nahezu alle Lebenslagen. Leistungsfähige und preiswerte Systeme, die sich unkompliziert und intuitiv bedienen lassen, finden große Verbreitung. Allerorts wird gescannt und digitalisiert: Bilder, Bücher, Karten, Landschaften, Räume, Museumsobjekte. Und wer fotografiert heutzutage noch mit Film? Auch im Bereich der Archäologie werden herkömmliche Methoden immer häufiger durch digitale Methoden abgelöst. Berichte, Pläne, Fotos usw. existieren oft nur noch als Datei. Viele dieser digitalen Daten sind es aufgrund ihrer Einzigartigkeit und hervorragenden Bedeutung wert, langfristig oder sogar dauerhaft bewahrt - archiviert - zu werden.

Nun ist allerdings die Informationstechnik heutzutage im Normalfall recht wenig auf Dauerhaftigkeit ausgerichtet. Ganz im Gegenteil bestimmen kürzeste Innovationszyklen das Geschehen und die neue Technik nimmt nur selten Rücksicht auf alte Daten. Wer kann schon heute noch Disketten lesen oder MODs? Wie lange werden CDs noch unterstützt werden? Niemand erwartet noch ernsthaft, Daten lesen zu können, die vor drei oder mehr Systemgenerationen geschrieben wurden. Wie soll man da etwas dauerhaft bewahren? Es ist zwar zu erwarten, dass die Archivierung digitaler Daten in der nächsten Zeit auch als Standard-IT-Thema stärker in den Vordergrund tritt und hoffentlich bald auch preiswerte kommerzielle Lösungen angeboten werden. Bis dahin bleibt allerdings nur, sich mit den Problemen vertraut zu machen und, Schritt für Schritt, nach individuellen Lösungen zu suchen. Mit Kenntnis einiger grundlegender Methoden kann man die Richtung vorgeben und erste Schritte gehen, ohne Datenverluste zu riskieren und auch ohne zukünftig möglichen Archivsystemen vorzugreifen.

Warum lassen sich digitale Daten eigentlich so schlecht aufheben? Zwei wesentliche Problembereiche lassen sich dafür identifizieren: Komplexität und Alterung.

Herkömmliche, analoge Dokumente sind in der Regel direkt und mit bloßem Auge lesbar und erkennbar. Dieses einfache Prinzip hat zu einer weitreichenden Verbreitung geführt und gewährleistet gleichzeitig, dass man Dokumente aus lang zurückliegender Zeit heute noch lesen kann, sofern das Medium erhalten geblieben ist.

Im digitalen Bereich finden wir eine andere Situation vor. Hier erschließt sich der Inhalt oder die Funktionalität nur über ein komplexes System aus Hardware, Software, Datenträger und Datenformat. Sobald sich in diesem komplexen System nach der Datenaufzeichnung etwas ändert, kann man nicht mehr gewiss sein, auf den ursprünglichen Inhalt bzw. die ursprüngliche Funktionalität noch zurückgreifen zu können. Ohne eine passende Systemumgebung ist der Datenträger mit all seinen Daten wertlos.

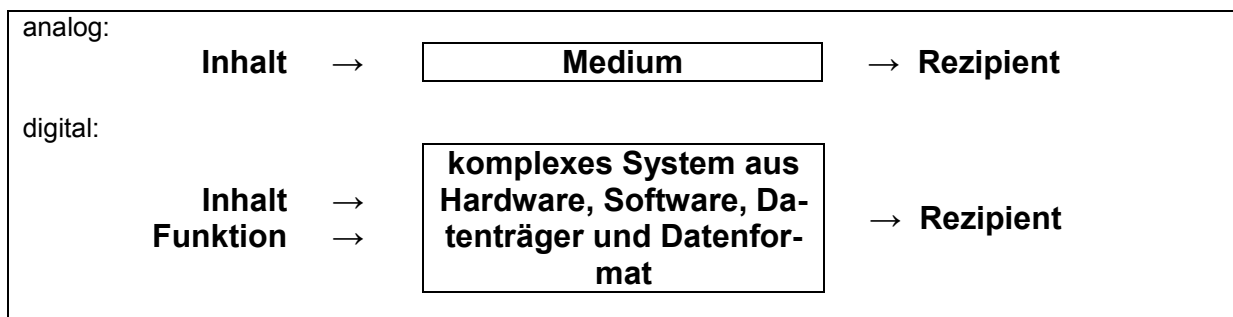


Abb. Komplexität eines analogen Systems (oben) und eines digitalen Systems (unten)

Die Alterung betrifft analoge Dokumente ebenso wie den digitalen Bereich. Allerdings laufen Alterungsprozesse bei analogen Dokumenten meist langsam und kontinuierlich ab, so dass man sie gut vorhersehen und Vorsorge treffen kann. Im digitalen Bereich sind Alterungsprozesse dagegen schnell und abrupt, was im Zusammenhang mit der Systemkomplexität eine Vorhersage äußerst schwierig macht. Das liegt daran, dass digitale Informationstechnologien auf enormen Speicherdichten basieren und nur wenig Platz für Redundanzen lassen. Digitale Daten sind also äußerst fehleranfällig. Hinzu kommen äußerst kurze technische Innovationszyklen von nur wenigen Jahren, wobei Nachhaltigkeit und Bewahrung bisher keine wesentliche Rolle spielen, meist zählt nur das Tagesgeschäft.

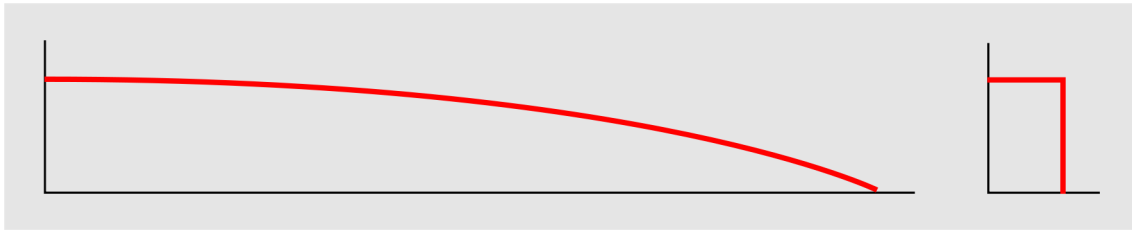


Abb. Alterungsprozess: links analog (kontinuierlich, langlebig) und rechts digital (Abrupt, kurzlebig)

Digitale Daten haben aber auch einen entscheidenden Vorteil, sie lassen sich nämlich ohne Informationsverlust kopieren und damit immer wieder erneuern. Dies ist bei analogen Medien nicht möglich. Ein gealtertes, verblasstes Foto hat unwiederbringlich Informationen verloren und man muss mit ggf. äußerst aufwändigen Methoden für den Erhalt wenigstens der verbliebenen Informationen sorgen, wenn schon das Medium nicht zu retten ist. Dazu bedient man sich dann übrigens gern der Digitalisierung.

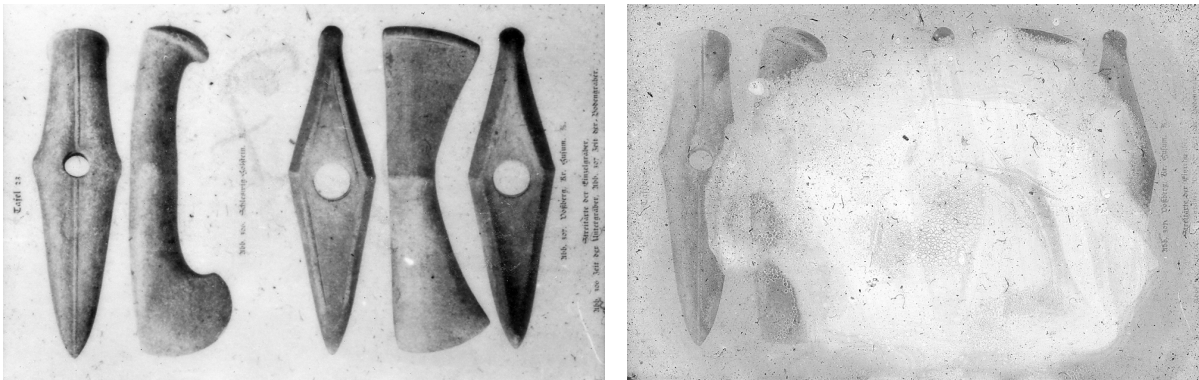


Abb. Alterungsprozess analoger Medien: links Kontaktabzug auf Papier (Xerokopie, 1990er Jahre) und rechts Diapositiv (2011)

Der Verband der Landesarchäologen Deutschlands erkannte die bestehenden Probleme und regte auf dem Kolloquium des Jahres 2008 die Gründung der AG „Archivierung“ innerhalb der Kommission „Archäologie und Informationssysteme“ an. Diese AG hat sich zum Ziel gesetzt, allgemein verfügbare methodische und technische Ansätze zur Archivierung digitaler Daten zu begutachten sowie praxisorientierte Empfehlungen, Vorschläge und Tipps zu geben. Durch eine Analyse der Situation in den archäologischen Landesbehörden Deutschlands wurden wichtige Themenbereiche mit dringendem Handlungsbedarf identifiziert. Mit dem vorliegenden Ratgeber soll diesen Zielen im Rahmen der bestehenden Möglichkeiten entsprochen werden.

## 2. Grundsätze

(Reiner Göldner)

Archivierung ist nicht immer gleich Archivierung. Der Begriff „Archivierung“ wird im traditionellen Archivwesen anders gebraucht, als es im modernen IT-Bereich oft üblich ist. Während „Archivierung“ traditionell mit Dauerhaftigkeit einhergeht, gibt man sich im IT-Bereich oft schon mit einigen Jahren Haltbarkeit zufrieden. In diesem Ratgeber wird jedoch auch hinsichtlich der digitalen Daten ausdrücklich die archivarische Sichtweise vertreten, die von einer dauerhaften, also unbegrenzten Erhaltung ausgeht. Im Vergleich zur herkömmlichen Archivierung geht es jedoch nicht nur um Dokumente, sondern um ggf. komplex strukturierte digitale Inhalte oder sogar um Funktionalitäten.

**Das Ziel der digitalen Archivierung ist es, digitale Inhalte oder digitale Funktionalitäten dauerhaft zu erhalten und verfügbar zu machen.**

Einige wichtige Anforderungen sollen hier vorab schon einmal erwähnt werden. Der tägliche Grundschutz der IT-Systeme und Daten soll einerseits Vertraulichkeit und Integrität sichern und andererseits für Verfügbarkeit sorgen, also dem Verlust vorbeugen. Die Gefährdungen sind vielfältig und entsprechend umfangreich die zu treffenden Gegenmaßnahmen. Passwörter, Virenschutz und Backups seien hier nur beispielhaft genannt, ausführliche Informationen erhält man auf den Seiten des Bundesamtes für Sicherheit in der Informationstechnik (BSI). Die Umsetzung des IT-Grundschutzes ist eine unbedingte Grundlage der Archivierung und erfordert umfangreiche Maßnahmen mit erheblichen Aufwänden, deren Umsetzung hier allerdings als gegeben vorausgesetzt wird. In diesem Ratgeber werden also im Wesentlichen nur die Themen angeschnitten, die über die tägliche IT-Grundsicherung hinausgehen.

Weiterhin spielen Integrität, Authentizität, ggf. Rechtsverbindlichkeit sowie Fragen der Zugänglichkeit (z.B. Urheberrecht, Persönlichkeitsschutz), und des Datenschutzes im Bereich der Archivierung eine wichtige Rolle (vgl. Abschnitt 4.4).

Zur Archivierung kommen beispielsweise die Methoden Hardcopy, Erhaltung, Migration und Emulation in Frage, sie werden im Abschnitt 6 näher erläutert. Als wegweisender Standard ist OAIS (ISO 14721) inzwischen weit verbreitet und anerkannt (vgl. Abschnitt 10).

Im fachlichen Bereich kommt der Auswahl archivwürdiger Dokumente und Daten eine wichtige Rolle zu. Dabei ist abzustimmen, welche Daten den laufenden Archivierungsaufwand rechtfertigen (vgl. Abschnitt 4.1).

### 3. Organisation der Archivierung

(Reiner Göldner)

Bei der Beschreibung der Methoden zur Archivierung digitaler Daten wurde schon deutlich, dass im digitalen Bereich sehr viel stärker als im herkömmlichen analogen Bereich gilt, dass Archivierung eine regelmäßige Betreuung erfordert. Diese Betreuung ist eine komplexe Aufgabenstellung und kann nicht nebenbei erledigt werden. Im Idealfall ist das digitale Archiv eine betriebliche Organisationseinheit, wie jedes andere Archiv auch (Aktenarchiv, Bildarchiv, Kartenarchiv, Bibliothek, Fundarchiv, ...). Es erfordert starke Kompetenzen aus den Bereichen Informationstechnik und Archivierung sowie entsprechende personelle und finanzielle Ressourcen.

Digitale Archivierung darf nicht verwechselt werden mit dem Thema Datensicherheit. Während letzteres zum IT-Tagesgeschäft gehört und nur der unmittelbaren Verfügbarkeit von Daten dient, gehen die Aufgaben der digitalen Archivierung weit über das Repertoire des IT-Ressorts hinaus und können von diesem nicht allein getragen werden!

Komplexe Aufgaben benötigen eine sorgfältige Vorbereitung und Planung. Folgende Punkte spielen dabei eine wichtige Rolle:

- die Verantwortlichkeit für das digitale Archiv muss festgelegt werden,
- es müssen finanzielle und personelle Ressourcen zugewiesen werden,
- es muss formell bestimmt und beschrieben werden, was archiviert werden soll,
- der technische Rahmen mit Erhaltungsmethoden, Datenformaten usw. muss abgesteckt werden,
- es sind Abläufe und Termine zu organisieren und
- die Ergebnisse sind einem Qualitätsmanagement zu unterziehen.

Abseits der großen Archive, die Archivierung als Kernaufgabe betreiben und damit auch der Archivierung digitaler Daten schon erhebliche Aufmerksamkeit widmen, gibt es eine ganze Anzahl von Institutionen, die aus unterschiedlichen Gründen ihre digitalen Unterlagen und Dokumente selbst archivieren wollen oder müssen. Typisch dafür sind Museen oder Fachämter, deren Dokumentation zunehmend digital erfolgt. Hier werden unwiederbringliche Informationen erarbeitet, die zweifellos archiviert werden müssen, nur kommt eine Auslagerung in ein „großes“ Archiv in der Regel nicht in Frage, weil der Informationsbestand „lebendig“ ist, sich also in steter Bearbeitung und Benutzung befindet.

Für solche kleineren Archive sind Archivierungslösungen erforderlich, die als kleine Bausteine und möglichst ohne große Kosten in die tägliche Arbeit einbezogen werden können. Solche Bausteine lassen sich an bestimmten Stellen aus den „großen“ Konzepten und Pilotprojekten isolieren. Diese Bausteine ermöglichen es, den Weg zur Archivierung digitaler Daten „Schritt für Schritt“ zu gehen.

Ein entscheidender Schritt ist es, die zu verwendende(n) Archivierungsmethode(n) festzulegen. Diese Entscheidung trifft man anhand der Eigenschaften und Funktionen der zu archivierenden Daten. Im Anschluss sind die wesentlichen Parameter dieser Methode(n) einzugrenzen, also z.B. die Datenformate für die Migrationsmethode. Damit ergeben sich die wesentlichen Trittsteine für die weitere Arbeit. In vielen Fällen wird man damit ein gutes Stück vorwärts kommen, ohne gleich ganz großen Aufwand betreiben zu müssen. Man kann z.B. dafür sorgen, dass schon bei der Datenerzeugung bzw. Digitalisierung archivierungstaugliche Formate benutzt werden. Zumindest wächst damit die Menge der problematischen Daten erst einmal nicht mehr (so stark) an.

Seit November 2011 liegt die DIN 31645 „Information und Dokumentation - Leitfaden zur Informationsübernahme in digitale Langzeitarchive“ vor. Auf der Webseite des DIN ([www.din.de](http://www.din.de)) findet man dazu: „Diese Norm gibt organisatorische Leitlinien für die Übernahme von Informationen aus einem Informationssystem in ein Langzeitarchiv unter Wahrung der Datenintegrität und Authentizität. Anfallende Aufgaben für den Produzenten und das digitale Langzeitarchiv werden in Planungsschritten benannt und in allgemeiner Form beschrieben. Der Geltungsbereich der Norm umfasst Vorhaben, in denen Informationsübernahmen in ein digitales Langzeitarchiv geplant und durchgeführt werden. Sie bezieht sich sowohl auf einmalige Informationsübernahmen als auch auf den Aufbau von Prozessen zur wiederholten oder automatisierten Informationsübernahme.“

Vermutlich wird aber auf Dauer nur eine umfassende Strategie für die Archivierung erfolgreich sein. Es bleibt zu hoffen, dass die erforderlichen Bausteine demnächst zu akzeptablen Preisen verfügbar sind.



## 4. Archivierungsrichtlinien

### 4.1. Archivwürdigkeit

(Reiner Göldner)

Es wird in den meisten Fällen nicht erforderlich sein, alle Daten zu archivieren. Dies ist in traditionellen Archiven nicht üblich und sollte auch in digitalen Archiven nicht angestrebt werden. Viele Dateien rechtfertigen den für die Archivierung erforderlichen Aufwand nicht! Auch sollten die wissenschaftlichen Rohdaten ein ausgewogenes Verhältnis zu den Ergebnissen aufweisen. Prüfen Sie also, ob die Rohdaten bedeutsam und archivwürdig sind und treffen Sie eine sorgfältige Auswahl. Je weniger archiviert werden muss, desto sorgfältiger kann es erledigt werden. Im Zweifel geben die Archivgesetze Beistand. Eine gute Idee<sup>1</sup> ist es, die Dokumente und Daten daraufhin abzuklopfen, ob sie denn ihrerseits dem kulturellen Erbe zuzurechnen sind.

*Zur Ergänzung dieses Abschnitts sind noch die folgenden Themen von Interesse:*

- *Wie, nach welchen Kriterien, sollen die Daten ausgewählt werden? (Inhalt, Qualität, Kosten, ...)*
- *Wann erfolgt die Auswahl und durch wen? (Verantwortliche, **Verursacherprinzip**, Strukturvorgabe, Duplikatvermeidung)*
- *Was archivieren (für immer)? Was aufbewahren (befristet, ereignisbezogen, nicht zeitbezogen)? → vgl. CAD/GIS*
- *Inwieweit sollen Primärdaten/Rohdaten archiviert werden?*
- *Wie unterscheiden sich Archäologisch-wissenschaftliche Daten von Verwaltungsdaten? (gesetzl. Fristen)*
- *Archivierungs-Aufwand als Grund für die Auswahl (falls Aufwand keine Rolle spielt, kann ja gern alles archiviert werden ;-)*
- *Kosten sind proportional zur Datenmenge, Kostenbewußtsein (Archiv als kostendeckend arbeitenden externen Dienstleister denken) (→ ADS)*
- *Hinweis: Datenerstellung (Grabungsstandards) auch schon an Archivierungsstandards anpassen (→ DFG)*

### 4.2. Qualitätsmanagement

*Für diesen Abschnitt sind die folgenden Themen von Interesse:*

- *Wie ist die Dokumentation (inhaltlich) in 100 Jahren nutzbar? Ist sie das überhaupt? Wie sind die Erfahrungen mit 50 Jahre alter Dokumentation?*
- *Warum QM? → Garant der Haltbarkeit, dauerhaften Verfügbarkeit, Voraussetzung für Archivierbarkeit*
- *Inhaltliches QM ist Sache des Wissenschaftlers (nicht Archiv), formales QM (Formate, Metadaten, ...) ist Sache des Archivs.*
- *Hinweis auf ISO 9001*
- *Für QM sind organisatorische Regelung erforderlich.*
- *Bestimmte Qualitätsstandards verursachen ggf. erhöhte Aufwendungen schon bei der Erfassung.*
- *Was bedeutet: Qualitätsmanagement für digitale archäologische Daten?*

### 4.3. Schutz der Informationen

(Reiner Göldner)

Der Schutz der Informationen umfasst die Themenbereiche Vertraulichkeit, Integrität, Authentizität und Verbindlichkeit, die bei der Archivierung digitaler Daten ebenso wie bei der herkömmlichen Archivierung eine wichtige Rolle spielen:

- Vertraulichkeit (Zugriffsschutz) gewährleistet, dass nur berechnigte Personen die entsprechenden Daten lesen und Informationen über den Inhalt erlangen können.

---

<sup>1</sup> Quelle: M. Aufleger, LVR-Amt für Bodendenkmalpflege Rheinland



- Unter Integrität (Änderungsschutz) von Daten versteht man, dass diese vollständig und unverändert bleiben. Entscheidend ist dabei, dass man Verletzungen der Integrität erkennen (und ggf. beheben) kann.
- Authentizität (Fälschungsschutz) bedeutet, dass eine Nachricht oder Information zuverlässig und überprüfbar einem Absender bzw. Urheber zugeordnet werden kann.
- Verbindlichkeit heißt, dass sich Inhalt und Urheberschaft einer Nachricht bzw. Information nicht abstreiten lässt und ggf. auch rechtsverbindlich nachweisbar ist (Revisionssicherheit, Rechtsverbindlichkeit).

Einfache Methoden zum Schutz von Informationen beruhen darauf, dass bestimmte individuelle Merkmale geprüft werden. Änderungsdatum, Dateigröße und Prüfsummen können zur Überprüfung von Integrität genutzt werden. Authentizität kann über Passwörter, Schlüssel, PIN o.ä. überprüft werden (und neuerdings auch über Fingerabdrücke und Iris-Scans).

Wichtige und verlässliche Verfahren zum Schutz von Informationen gegen unbefugtes Lesen und Verändern stellt die Kryptographie zur Verfügung (Verschlüsselung). Über Zertifikate kann die Integrität und Authentizität digitaler Daten geprüft werden. Dazu dienen z.B. Public-Key-Verfahren mit privaten und öffentlichen Schlüsseln. Solche elektronischen Signaturen dienen dem gleichen Zweck wie Unterschriften auf Papierdokumenten. In Deutschland gelten diesbezüglich das Signaturgesetz (SigG) und die Signaturverordnung (SigV), wobei einfache, fortgeschrittene und qualifizierte elektronische Signaturen unterschieden werden.

Je nach Anforderungen an den Schutz der Informationen muss entschieden werden, welche Methoden zum Einsatz kommen sollen. Integrität und Authentizität sind dabei einfacher realisierbar als Rechtssicherheit, eine interne Zertifizierung eher umsetzbar als eine externe Zertifizierung durch ein Trust Center. Im Zweifelsfall wird es unerlässlich sein, entsprechende Experten einzubinden.

Jede Datenmigration ist ein Eingriff in die Integrität der Daten. Wichtig dabei ist aber, dass der Inhalt erhalten bleibt (das, was der Autor bzw. Urheber übermitteln wollte). Dazu muss für jedes Dokument (oder zumindest für jede Klasse von Dokumenten) festgelegt werden, welche Eigenschaften signifikant sind und bei der Migration unbedingt zu erhalten sind (→ Metadaten). Einige Überlegungen dazu findet man im Aufsatz „Archäologische digitale Daten – Authentizität und Funktionalität“ (Bibby/Göldner).

## 5. Metadaten

(Reiner Göldner)

Metadaten enthalten Informationen über das archivierte digitale Dokument. Mit ihnen wird in erster Linie das Suchen und Finden unterstützt. Aber auch die internen Archivierungsaufgaben erfordern spezielle Informationen, die in den Metadaten unbedingt enthalten sein sollten. Je detaillierter und vollständiger zum Beispiel das Datenformat eines Dokuments beschrieben ist, desto einfacher lässt sich ggf. seine Migration in ein aktuelleres Format organisieren und durchführen. Bei den Metadaten kann man inhaltliche, technische und organisatorische Aspekte unterscheiden. Ein wichtiges Kriterium für die Auswahl von Metadaten ist die zukünftige Nutzung.

Inhaltliche Metadaten kann man beispielsweise im weit verbreiteten Dublin-Core-Format (DC) ablegen. Die Kernelemente dieses Formats sind im nachfolgenden Abschnitt beschrieben.

Für Technische Metadaten kann man sich am „PREMIS Data Dictionary for Preservation Metadata“ orientieren, welches ein internationaler Metadaten-Standard zur Unterstützung der Langzeitarchivierung digitaler Objekte ist. Technische Metadaten enthalten alle wesentlichen technischen Dateimerkmale, die für eine Präsentation des Inhalts erforderlich sind. Im Zentrum stehen dabei detaillierte Angaben zum Dateiformat und zur Erzeugersoftware (s. Abschnitt 5.2).

Organisatorische Metadaten beinhalten z.B. Zugangsregeln, Nutzungsrechte und Archivierungsfristen.

### 5.1. Dublin-Core-Standard

DC-Tag	dt. Begriff	Beschreibung
identifier	ID	eindeutiger Identifikator innerhalb eines bestimmten Kontextes
format	Formatangabe	Angabe zum Dateiformat
type	Daten-Typ	Auswahl aus: Bild, Text, interaktiv, Software, Dataset, Kollektion, ...
language	Sprache	z.B.: de, en, fr, es, it, la, ... (ISO 639)
title	Titel	Titel des Dokuments, Name der Ressource
subject	Thema	Schlagworte, die den Inhalt beschreiben
coverage	Geltungsbereich	Eingrenzung des Inhalts (z.B. räumlich, zeitlich)
description	Beschreibung	Kurzzusammenfassung (Abstract) zum Inhalt, ggf. Anmerkungen
creator	Urheber	verantwortlicher Urheber oder Verfasser
publisher	Verleger	veröffentlichende Instanz, typischerweise Verleger, Herausgeber
contributor	Mitwirkende	weitere Personen mit Beitrag zur Erstellung der Ressource
rights	Rechte	Angaben zu Rechten bzgl. der Ressource
source	Quelle	Angabe zu Quellen, von denen die Ressource abgeleitet wurde
relation	Verweis	Verweis auf andere Ressourcen
date	Datum	charakteristisches Datum bzgl. des Dokuments

Für archäologische Zwecke muss DC ggf. durch fachspezifische Angaben erweitert werden. Anregung dazu kann der Archäologische Datenaustausch-Standard ADeX geben. Außerdem lohnt es sich, über die Anbindung bestehender Fachinformationssysteme nachzudenken und dort etablierte Identifikatoren (Schlüsselfelder) mit einzubeziehen.

### 5.2. Metadaten für Archivierungsaufgaben

(Sigmar Fitting)

Ein wesentlicher Aspekt bei der Auswahl der zu archivierenden Dateien stellt die Dokumentation der verwendeten Formate und die Validierung der Dateien gemäß der standardisierten Spezifikationen.

Ein dazu geeignetes Tool stellt etwa JHOVE dar. JHOVE prüft Dateien gegen eine Liste von Formatspezifikationen, so dass man sehr spezifische Formatinformationen erhält. Details findet man unter <http://hul.harvard.edu/jhove/> bzw.

[http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor\\_handbuch\\_artikel\\_313.pdf](http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_313.pdf).



Zum Testen ist die von der Harvard University entwickelte Version 1.x am besten geeignet, da sie auch über ein Testtool mit GUI verfügt. Im Moment wird von verschiedenen Organisationen an JHOVE 2.x gearbeitet (<https://bitbucket.org/jhove2/main/wiki/Home>).

*„Prinzipiell kann man nie genug über eine archivierte Datei wissen, jedoch kann es durchaus sinnvoll sein, extrahierte Metadaten einmal auf ihre Qualität zu überprüfen und gegebenenfalls für die Langzeitarchivierung nur indirekt relevante Daten herauszufiltern, um das Archivierungssystem nicht mit unnötigen Daten zu belasten. Beispiel für ein solches Tool ist „JHOVE“ (das JSTOR/Harvard Object Validation Environment der Harvard University Library, <http://bul.harvard.edu/jhove/>), mit dem sich auch Formaterkennung und Validierung durchführen lassen. Das Tool ist in Java geschrieben und lässt sich auch als Programmier-Bibliothek in eigene Anwendungen einbinden. Die generierten technischen Metadaten lassen sich sowohl in Standard-Textform, als auch in XML mit definiertem XML-Schema ausgeben.*

...

*Validierungstools für Dateiformate stellen sicher, dass eine Datei, welche in einem fraglichen Format vorliegt, dessen Spezifikation auch vollkommen entspricht. Dies ist eine wichtige Voraussetzung für die Archivierung und die spätere Verwertung, Anwendung und Migration beziehungsweise Emulation dieser Datei. Das bereits erwähnte Tool „JHOVE“ kann in der aktuellen Version 1.1e die ihm bekannten Dateiformate validieren; verlässliche Validatoren existieren aber nicht für alle Dateiformate. Weit verbreitet und gut nutzbar sind beispielsweise XML Validatoren, die auch in XML Editoren wie „oXygen“ (SyncRO Soft Ltd., <http://www.oxygenxml.com>) oder „XMLSpy“ (Altova GmbH, <http://www.altova.com/XMLSpy>) integriert sein können.“*

Auszug aus dem Nestor-Handbuch Version 2, Kap. 7, S. 16,17

Eine sehr viel gröbere Erkennung bietet DROID auf Basis der „PRONOM Format Registry“ der National Archives of Great Britain. Hierzu findet man Details unter <http://sourceforge.net/projects/droid/files/droid/> bzw. <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx> .

### 5.3. Dokumentenmanagement

(Reiner Göldner)

Archive müsse gewährleisten, dass man ihren Inhalt finden kann. Zum Suchen und Finden sind Metainformationen über den Inhalt der archivierten Daten erforderlich. Dazu gehören auch Rechte, Fristen, Nachweise, ggf. Signaturen usw. Digitale Archive müssen aber auch in kurzen Fristen überprüft und überarbeitet, die Daten migriert werden. Dies wird per Hand kaum möglich sein, daher ist es erforderlich, die zur Organisation und (automatisierten) Durchführung der Erhaltungsmaßnahmen notwendigen Informationen vorzuhalten. Solche Metainformationen lassen sich am besten in einem Dokumentenmanagementsystem verwalten.

Zur Ergänzung dieses Abschnitts sind noch die folgenden Themen von Interesse:

- *Dokumentenmanagementsystem oder Datenträgerverwaltung: Welche Systeme gibt es?*
- *Alternative: strukturiertes Dateisystem*
- *Weitere Alternativen: formloses Inhaltsverzeichnis hilft schon manchmal, ggf. Tabellen*
- *Dokumentenmanagement für Geschäftsprozesse (befristete Bewahrung) und Wissenschaftliche Dokumentation (Archivierung) gleichermaßen nutzbar (Synergie)*
- *DMS bildet auch Geschäftsprozesse ab, das ist in der Archivierung nicht erforderlich*



## 6. Archivierungsmethoden

(Reiner Göldner)

Wenn Archivierung gewährleisten soll, dass digitale Inhalte und Funktionalitäten dauerhaft bewahrt und zur Benutzung erhalten werden, müssen also Maßnahmen gegen den Einfluss der Alterung ergriffen werden. Dies ist integrale Aufgabe der Archivierung in analogen ebenso wie in digitalen Archiven. Nur kann man eben bei Papier viele Jahre abwarten, während man bei magnetischen Datenträgern oft schon nach einigen Monaten wieder aktiv werden muss.

Nachfolgend werden die wichtigsten Methoden zum Erhalt der Benutzbarkeit im Bereich digitaler Daten aufgeführt und kurz beschrieben. Leider ist keine universelle Methode darunter, die alle Probleme lösen kann, so dass man in einem digitalen Archiv ggf. mehrere Methoden nebeneinander betreiben muss.

### 6.1. Hardcopy

Mit „Hardcopy“ ist meist der Ausdruck eines digitalen Dokuments auf Papier gemeint oder, etwas allgemeiner, die Ausgabe auf ein analoges Medium. Für die Langzeitarchivierung ist dabei die relativ lange Lebensdauer der benutzten Medien bedeutsam. Zum Beispiel geht man bei säurefreiem Papier von einer Haltbarkeit von über 200 Jahren aus. Für spezielle Medien, z.B. LE500 Papier oder LE500 Mikrofilm (PET) wird eine ANSI-zertifizierte Lebenserwartung von 500 Jahren angegeben. Diese im Vergleich zu digitalen Medien extrem lange Haltbarkeit führt zu einem minimalen Betreuungs- und Erhaltungsaufwand. Die Medien können natürlich auch nicht einfach weggeschlossen werden, sondern es muss z.B. für Klimatisierung und Brandschutz gesorgt werden, aber dies erledigt ja im Zweifelsfall ein herkömmliches Archiv.

Vorteilhaft ist weiterhin, dass man die Dokumente direkt und einfach wieder lesen kann. Selbst wenn man in Betracht zieht, dass man Mikrofilme stark vergrößern muss, so ist die dazu notwendige Technologie doch wesentlich einfacher als die Computertechnologie.

Hardcopies haben jedoch auch entscheidende Nachteile. Digitale Dokumente beinhalten oft schon eine Reihe von Informationen, die sich schlecht oder gar nicht ausdrucken lassen. Man denke nur an Ebenenstrukturen, animierte Grafiken oder interaktive Elemente wie Hyperlinks. Bei Plänen bzw. Karten im CAD- oder GIS-Bereich erreicht man oft nur bei Auswahl bestimmter Ebenen eine übersichtliche Druckansicht, dabei geht die Flexibilität der Ebenenstruktur verloren. Außerdem gibt es oft eine Reihe von Attributen zu den einzelnen Zeichnungselementen, die nur interaktiv sinnvoll nutzbar sind und beim Ausdruck größtenteils verloren gehen.

Funktionalitäten, also Algorithmen und Programme, lassen sich per Hardcopy nur beschreibend abbilden, sie funktionieren also nicht auf dem Papier. Es gibt zwar Ansätze, digitale Daten maschinenlesbar auf Papier oder Film zu bringen (z.B. Barcodes), sie haben aber für die Archivierung bisher nur geringe Bedeutung erlangt, da davon nur die Haltbarkeit des Datenträgers betroffen ist, das gesamte restliche System aber außer Acht gelassen wird.

Bei Anwendung der Hardcopy-Methode sollte man besonderes Augenmerk auf die Art der Visualisierung legen. Möglicherweise ergibt sich ein deutlicher Mehrwert, wenn man verschiedene Visualisierungen nebeneinander legt. Im CAD- oder GIS-Bereich kann man z.B. Pläne bzw. Karten in Verbindung mit textlichen Darstellungen von Attributen oder Layerstrukturen ablegen, so dass diese Informationen nicht verloren gehen. Es können auch ergänzende Codebestandteile angegeben werden.

Hardcopies sollten so angelegt werden, dass sie maschinenlesbar sind. Das bedeutet im einfachsten Fall dass sie ebenso hochauflösend gescannt werden können wie sie geschrieben sind. Hilfreicher ist es jedoch, wenn Metadaten integral unterstützt werden. Maschinenlesbarkeit kann z.B. durch Text (OCR) oder Strichcodes gewährleistet werden.

Die Hardcopy-Methode eignet sich gut als Ergänzung in Fällen, in denen andere Archivierungsmethoden problematisch sind.



## 6.2. Erhaltung

Erhaltung bedeutet, dass man die Lebensdauer des Gesamtsystems maximiert. Man versucht, durch Reduzierung von Umwelteinflüssen und optimale Pflege das System möglichst lange betriebsbereit zu halten. Solange dies gelingt, werden die digitalen Daten, die ja Bestandteil dieses Systems sind, verfügbar und nutzbar sein.

Der Vorteil dieser Methode ist, dass hier nicht verändernd in das komplexe System aus Hardware, Software, Datenträger und Datenformat eingegriffen wird. Wenn bestimmte Daten außerhalb ihrer ursprünglichen Systemumgebung entweder gar nicht oder nicht authentisch wiedergegeben werden können, schließt das andere Archivierungsmethoden meist aus. Außerdem bleiben komplexe Dokumente und Funktionalitäten problemlos benutzbar.

Es ist leicht einsehbar, dass diese Methode nur für einen beschränkten Zeitraum anwendbar sein wird, da sie den Alterungsprozess nur verlangsamen kann, nicht aber verhindern. Die Methode eignet sich also streng genommen nicht für eine dauerhafte Bewahrung. Außerdem müssen ja quasi alle benutzten Systeme erhalten werden, zumindest als Prototyp, weil alle Daten eben ihre originale Systemumgebung benötigen. Dies ist sehr aufwändig.

Die Erhaltungs-Methode kann gut durch den Begriff „Computermuseum“ verdeutlicht werden. Es gibt auch tatsächlich solche Museen, die sich der Bewahrung alter Computersysteme verschrieben haben und auch das eine oder andere System am Leben erhalten.

## 6.3. Migration

Migration bedeutet, dass Datenträger und Datenformat regelmäßig aktualisiert und damit an die Erneuerungen der restlichen Systemkomponenten angepasst werden. Damit hat man insgesamt gesehen immer ein aktuelles System. Vorteilhaft dabei ist, dass man weder alte Hardware und Software sammeln (Erhaltung) noch Nachbildungen dieser Systeme erstellen (Emulation) muss. Informationen und Funktionalitäten sind immer direkt auf aktuellen Systemen verfügbar und man muss lediglich für deren Betrieb sorgen. Außerdem sind die Daten auf einer breiten Palette moderner PCs direkt lesbar und können somit problemlos breit verfügbar gemacht werden.

Bei einer Aktualisierung des Systems muss man jedoch neben Hardware und Software ggf. auch für eine Aktualisierung der Daten sorgen. In einigen Fällen wird es auch notwendig sein, veraltete Datenformate zu ersetzen. Beides ist zumeist technisch unkritisch, aber es muss eben für den gesamten Archivbestand organisiert und regelmäßig durchgeführt werden. Und je älter die Daten sind, umso schwieriger wird die Migration in ein aktuelles Format. Ein prinzipieller Nachteil dabei ist, dass jede Aktualisierung des Datenformats einen Eingriff in Integrität und Authentizität der Daten bedeutet, was ggf. auch Auswirkungen auf Revisionsicherheit und Rechtsverbindlichkeit hat. Die konkreten Migrationsschritte müssen daher sorgfältig konzipiert und überprüft werden. Da gegenwärtig kaum Werkzeuge dazu angeboten werden bleibt oft nur die manuelle Kontrolle.

Der Auswahl von bestimmten Datenformaten kommt bei der Migration eine erhöhte Bedeutung zu. Um Probleme zu reduzieren, sollten möglichst langlebige, offene und standardisierte Formate Verwendung finden, die bei der Datenübernahme ins Archiv verifiziert werden müssen. Um den Migrationsaufwand in Grenzen zu halten, sollte man sich auf einige wenige Formate beschränken und darauf achten, dass sich die Daten automatisiert migrieren lassen. Dazu sind ausreichende technische Metadaten vorzuhalten. Die Migrations-Methode ist jedoch prinzipiell nicht an bestimmte Formate gebunden, so dass die Formatfrage in jedem Fall sekundär ist und immer im Kontext der Methode betrachtet werden muss.

Das gleiche gilt für die Datenträger. Auch hier ist Langlebigkeit natürlich von Vorteil, aber methodisch nicht ausschlaggebend. Viel wichtiger ist, dass nach der Bewahrungsphase problemarm migriert werden kann, was man ggf. mit kurzlebigeren Datenträgern besser realisieren kann als mit langlebigeren. In vielen Fällen wird man daher eine Auslagerung auf externe Datenträger vermeiden.

#### 6.4. Emulation

Emulation bedeutet, dass ein älteres Computersystem auf einem aktuellen System funktionell nachgebildet wird. Das emulierte System verhält sich dabei im Idealfall genau so wie sein Original und kann demnach auch Originaldaten nutzen. Typische Beispiele dafür sind die Windows-Emulationen für Macintosh- oder LINUX-Systeme und umgekehrt sowie eine Vielzahl von Emulationen für ältere Homecomputer (Atari, C64 usw.).

Ein Vorteil dieses Verfahrens liegt darin, dass die Originaldaten unverändert verwendet werden können. Die Daten werden nur binär lesbar gehalten und ggf. auf einen aktuellen Datenträger kopiert. Dabei bleibt das Datenformat unangetastet und die Daten können originalgetreu verwendet werden. Außerdem ergibt sich damit eine einfache Möglichkeit Funktionalitäten zu bewahren. Das emulierte System kann nicht nur originale Dokumente lesen sondern auch originale Programme ausführen. Weit verbreitet sind z.B. Emulationen alter Homecomputer wie C64 und ATARI, die für fast alle aktuellen Systeme erhältlich sind und es ermöglichen, die alten Computerspiele noch zu nutzen.

Nachteilig wirkt sich aus, dass solche Nachbildungen durchaus aufwändig sind und dass sie oft keine vollständige Kompatibilität mit dem Original erreichen. Oft wird die volle Kompatibilität jedoch nicht erreicht, z.B. im Bereich der Peripherie. Ebenso wie bei der Erhaltungs-Methode müssen alle Systemvarianten emuliert werden, zu denen Daten existieren. Ist dies auch einfacher als der Erhalt der Hardware, so verbleibt doch ein sehr hoher Aufwand, wenn man diese virtuellen Systeme alle am Leben erhalten will.

#### 6.5. Sonderfall Datenbanken

Datenbanken sind eigentlich keine Archivierungsmethode, allenfalls ein Mittel zum Zweck. Sie weisen allerdings eine Vielzahl Eigenschaften auf, die sie für Langzeitarchivierung äußerst interessant machen.

Datenbanken, zumindest die etablierten Server-Varianten, sind stark auf Langzeitnutzung ausgelegt. Jedes Systemupdate beinhaltet stets Mittel und Methoden zur Migration und zum Erhalt der Datenintegrität (Welches Programm bietet dies sonst?). Damit wird die Migrationsmethode integral unterstützt. Weiterhin sind die Datenformate und die Grundfunktionalitäten stark standardisiert und weit verbreitet, so dass überdies ein breiter Austausch von Daten und Funktionalitäten, auch über Systemgrenzen hinweg möglich ist.

Natürlich gibt es auch hier Einschränkungen, weil die Datenbankmanagementsysteme und ihre internen Datenformate proprietär sind und der Zugriff nur über die entsprechenden Datenbankschnittstellen erfolgen kann. Außerdem sind komplexere Datenformate (z.B. Geodaten) ebenfalls meist proprietär und damit schlecht austauschbar. Diesen beiden Kritikpunkten muss man begegnen, wenn man z.B. Dokumente (Texte, Bilder usw.) direkt in Datenbanken archivieren möchte.

Innerhalb eines Archivsystems sind Datenbanken hervorragend für die Datenverwaltung geeignet, d.h. für die Bewahrung der deskriptiven Metadaten und Verwaltungsinformationen.

#### 6.6. Notfall und Einzelfall

Auch der Notfall ist natürlich keine Archivierungsmethode. Aber Not macht erfinderisch und es gibt noch Methoden, die es uns erlauben, alte Datenbestände wieder lesbar zu machen, wenn keine der bisher beschriebenen Methoden für die Benutzbarkeit gesorgt hat.

Wenn man hochwertige Daten auf einem alten Datenträger hat und kein adäquates System vorrätig hat bzw. findet (Museum), kann man Systemkomponenten nachbauen, die diesen Datenträger lesen können. Oder man baut Software nach, die ein altes Datenformat wieder lesen kann. Voraussetzung ist in jedem Fall, dass man eine ausreichende Beschreibung des Datenträgers (Funktion) bzw. des Datenformats zur Verfügung hat.

In manchen Fällen kann es sogar einfacher sein, eine solche Beschreibung über die Zeit zu retten, als die Komponente selbst. Dies wird vor allem bei speziellen seltenen oder gar einmaligen Datenforma-



ten der Fall sein. Als Beispiel seien hier XML-Dateien angegeben, die strukturierte Daten in Textform speichern. Während der Text an sich unproblematisch lesbar ist (und bleiben wird), erschließt sich die Struktur nur durch spezielle Software oder aber durch eine Beschreibung, die, ebenfalls in lesbarer Form archiviert, den einfachen Nachbau der Funktionalität der Originalsoftware erlaubt. In solchen besonderen Fällen wird die Notfallvariante also zur Methode.

Wenn es um einmalige Datenrettung geht, kann man sich auch an speziell ausgerüstete Labore wenden, welche einen Datenträger mit empfindlichen Sensoren direkt abtasten und anschließend aus den so gewonnenen Messwerten unter günstigen Bedingungen auch wieder Daten rekonstruieren können. Die Kosten der Datenrettung werden einen aber schnell zu anderen Archivierungsmethoden zurück führen.

### 6.7. Zusammenfassung der Methoden

Betrachtet man die vorgestellten Methoden, so kann man feststellen, dass jede für sich ihre Stärken und Schwächen hat, aber allein nicht ausreichend ist. Insgesamt gesehen bieten Sie jedoch ein gutes Instrumentarium, um die Anforderungen der Archivierung digitaler Daten umfassend zu erfüllen. In der folgenden Tabelle werden wichtige Vor- und Nachteile der vorgestellten Methoden noch einmal zusammengefasst.

	einfache Formate	Dokument mit Layout	komplexe Strukturen	Funktionalität	Sonstiges
Hardcopy	ja	ja	nein	nein	lange Lebensdauer
Erhaltung	ja	ja	ja	ja	aufwändig, nicht dauerhaft
Emulation	ja	ja	ja	ja	aufwändig
Migration	gut	möglich	möglich	nein	gut beherrschbar
Datenbank	hervorragend	schlecht	nur relational	nein	verfügbar, gut bewährt

**Die vorgestellten Erhaltungsmethoden zeigen, dass Archivierung eine kontinuierliche Betreuung erfordert. Dies wird in herkömmlichen Archiven nicht angezweifelt und muss auch im digitalen Bereich akzeptiert werden. Der Methodische Ansatz macht außerdem deutlich, dass die Wahl von Datenträger und Datenformat nicht von primärer Bedeutung ist. Nicht die Lebensdauer entscheidet über den Erhalt, sondern die Methodik.**

## 7. Datenträger und Datenformate

(Reiner Göldner)

Vielen Diskussionen zur Archivierung liegt die Idee zugrunde, möglichst langlebige und haltbare Datenträger zu verwenden. Demzufolge werden Haltbarkeitswerte und zeitliche Fehlerraten ausgiebig diskutiert. Dem Trend dieser Diskussionen soll hier nicht widersprochen werden, jedoch sei angemerkt, dass in Abhängigkeit von der Archivierungsmethode auch noch andere Anforderungen als nur Haltbarkeit eine Rolle spielen. Für die Durchführung einer Migration wird es beispielsweise nicht unerheblich sein, dass die Daten möglichst direkt und ohne manuelle Interaktionen zugänglich sind, ein Stapel CDs wäre dabei also kontraproduktiv. Die Migrations-Methode wird also durch einen online-Zugang zu den Daten stark unterstützt und man wird sich eher für ein Festplattensystem oder eine Tape-Library entscheiden. Bei der Erhaltungs-Methode sieht es dagegen völlig anders aus und man wird auch auf eine lange Lebensdauer des Datenträgers hohen Wert legen.

**Was bei den Datenträgern anklingt gilt ebenso für die Datenformate, man kann die Diskussion über Vor- und Nachteile nicht von den Archivierungsmethoden trennen. Es ist zwingend erforderlich, dass man sich über die zu benutzende Archivierungsmethode im Klaren ist, bevor man Datenträger und Datenformate auswählt!**

Die vorgestellten Methoden schränken die nutzbaren Datenformate nicht ein, dies ist prinzipiell auch nicht erforderlich. Trotzdem ist es vorteilhaft und vermindert den Aufwand, wenn man

- stabile Datenformate verwendet, die weit verbreitet, nicht proprietär und standardisiert sind und
- einfache Datenformate bzw. Datenstrukturen verwendet, die problemarm zu migrieren sind.

Bezüglich der Tauglichkeit zur Archivierung sollen drei Level unterschieden werden:

- **bevorzugt:** Diese Formate sind sehr unproblematisch und sollten wenn möglich benutzt werden.
- **akzeptiert:** Diese Formate sind möglicherweise problematisch, d.h., sie erfordern ggf. einen erhöhten Migrationsaufwand. Sie können benutzt werden, wenn die bevorzugten Formate nicht verfügbar oder ungeeignet sind.
- **Kandidaten:** Diese Formate sind problematisch hinsichtlich einiger weniger Anforderungen. Sie können in begründeten Einzelfällen in Abstimmung mit dem Archiv benutzt werden, wenn andere Formate nicht verfügbar oder ungeeignet sind. Die Formate sollen in ihrer Entwicklung (z.B. Stabilität, Verbreitung) verfolgt werden.

Details zu den Datenformaten sind in einer Anlage aufgelistet.

### 7.1. Texte

(Andreas Brunn, Reiner Göldner)

Unter „Texte“ sollen Dokumente verstanden werden, die überwiegend Schriftzeichen enthalten. Man könnte also auch „Schrift-Dokumente“ sagen. Für die Archivierung ist von Bedeutung, ob die Texte formatiert sind oder nicht. Unformatierte Texte bestehen nur aus den Zeichen eines Zeichensatzes. Im Gegensatz dazu enthalten formatierte Texte eine Vielzahl von Informationen zur Darstellung von Zeichen: fett, kursiv, Schriftgrad usw., zum Layout (Blattrand, Einzug, ...) sowie auch durchaus zeichnerfremde Elemente (Grafiken, Abbildungen,...).

Einfache Texte werden seit Beginn der Computertechnik unterstützt, sind nach wie vor weit verbreitet und können in der Regel auch über Systemgrenzen hinweg gut gelesen und verarbeitet werden. Es ist auch nicht zu erwarten, dass sich an dieser Situation grundsätzlich etwas ändert. Probleme können sich jedoch durch unterschiedliche Kodierungen bzw. Zeichensätze ergeben, vor allem, wenn diese nicht bekannt sind. Gut verbreitete und damit weitgehend unproblematische Zeichensätze sind z.B.:

- US-ASCII (keine dt. Sonderzeichen!)
- Unicode UTF-8 (oder UTF-16)
- ISO 8859-1(Latin-1 westeuropäisch).

Aus den meisten üblichen Textverarbeitungsprogrammen heraus lassen sich Dokumente im Textformat, auch unter expliziter Auswahl des Zeichensatzes, speichern.

Einfache Texte können unstrukturiert sein (.TXT) oder strukturiert (.XML, .CSV). Strukturierte Texte enthalten Zusatzinformationen, die lesbar in den Text integriert sind. Bei der Archivierung muss nun



sichergestellt werden, dass diese Zusatzinformationen interpretiert werden können, was eine zusätzliche Beschreibung der Bedeutung bzw. Struktur dieser Zusatzinformationen erfordert.

Übliche Textverarbeitungsprogramme erzeugen zumeist formatierte Texte. Auch wenn es einige weit verbreitete Formate gibt, so sind diese doch meist nur kurzlebig und daher schlecht zur Archivierung geeignet, da ständige Migrationen erforderlich würden. Abhilfe soll ein normiertes Format schaffen, die ISO-Norm 19005 beschreibt PDF/A explizit für Archivierungszwecke. Ausgehend vom PDF-Format der Fa. Adobe wurde hier von allzu modernen und kurzlebigen Features abgesehen, um langfristige Stabilität zu erreichen. PDF/A ist gegenwärtig das einzige Format für formatierte Texte, das weitgehend uneingeschränkt für die Archivierung empfohlen werden kann. Zu berücksichtigen sind allerdings mehrere Versionen von PDF/A, wobei die einfachen Versionen (z.B. PDF/A-1b) nur das Erscheinungsbild des Dokuments abbilden und eine Extraktion des Textes (z.B. copy-past) nicht erlauben.

Eine Archivierungsvariante ergibt sich, wenn man das Schrift-Dokument als Bild betrachtet und in einem für Bilder geeigneten Format archiviert (s. Abschnitt 7.2).

#### Formate für Texte

**bevorzugt:**

- **TXT** für einfache Texte
- **PDF/A** für formatierte Texte

**akzeptiert (ggf. problematisch):**

- ggf. **CSV** für Tabellen (zzgl. Strukturbeschreibung)
- ggf. **TIFF** (→ Bilddokumente) für formatierte Texte

**Kandidaten (problematisch, weiter prüfen):**

- ggf. **SGML, XML** (zzgl. Strukturbeschreibung)

Zusammenfassung: Text-Formate zur Archivierung

## 7.2. Bilder

(Sigmar Fitting, Reiner Göldner)

Bei Bildern sind die Dateiformate mit einer Vielzahl von Eigenschaften und Attributen verknüpft, die schon bei der Erzeugung der Bilddaten, also beim Scannen, Fotografieren usw. berücksichtigt werden müssen. Dazu gehören z.B. Auflösung, Farbtiefe, Farbraum und viele andere. Dazu gibt es im Anhang eine ausführliche Beschreibung des „Workflow“. Nachfolgend kann der Fokus daher auf die Formate beschränkt bleiben.

Die Menge der o.a. Eigenschaften und Attribute erfordert eine sorgfältige Dokumentation der Formatspezifikationen der archivierten Bilddokumente in den Metadaten. Je detaillierter diese Dokumentation ist, umso leichter lassen sich erforderliche Migrationen planen und durchführen (vgl. Abschnitt 5.2).

Bei der Archivierung von Bildern sollte, in Anbetracht sinkender Preise für Foto- und Speichertechnik, auch darüber nachgedacht werden, Bilder mehrfach redundant zu archivieren. So können die Nachteile einzelner Formate vermieden werden und man kann später bei der Nutzung das zweckmäßigste Format auswählen

Eine weitere sehr gute Möglichkeit, wichtige Aufnahmen risikofrei für die nächsten 50 – 100 Jahre aufzubewahren stellt das Ausbelichten auf Fotopapier und Aufbewahren in altmodischen Archivschränken dar. Genau so ernsthaft wird auch immer wieder die Ausbelichtung auf Dia-oder Mikrofilm diskutiert (vgl. Abschnitt 6.1)

#### RAW-Bildformate

Die besten Möglichkeiten, um möglichst alle „gemessenen“ Bildinformationen abzuspeichern, bieten die sogenannten RAW-Bilder (Rohdaten). Diese RAW-Dateien bieten die maximale Qualität bei ge-

ringst möglicher Speichergröße der verlustfrei komprimierten Daten allerdings zum Preis einer sehr geringen Kompatibilität.

Ein sehr hoffnungsvoller Ansatz bot 2004 die Einführung und Offenlegung des DNG-Formates als potentieller Standard für das „digitale Negativ“, dieser Standard konnte sich allerdings bis heute nicht so recht durchsetzen allerdings gibt es zum Beispiel mit Adobe Camera Raw ein Tool zum Konvertieren der proprietären RAW-Formate.

Nikon etwa hat bis heute noch nicht alle Spezifikationen des eigenen Raw-Formats offengelegt bei Canon verändern sich die Spezifikationen von einem Kameramodell zum Nächsten. Im HighEnd Bereich sieht die Sache nicht wirklich besser aus, hier bleibt die weitere Entwicklung abzuwarten und bis dahin (soweit alle Bearbeitungsmöglichkeiten erhalten bleiben sollen) die Kamera eigenen RAW-Dateien zu archivieren oder die Raw-Dateien in das DNG-Format zu konvergieren.

Im Scanbereich sieht es noch düsterer aus, etwa Imacon unterstützt in der neuesten Softwareversion das eigene „3f“-Format nicht mehr oder Vuescan-DNG's lassen sich nur mit Vuescan weiter bearbeiten, wenn man alle Bearbeitungsmöglichkeiten erhalten will.

Aufgrund der fehlenden Kompatibilität können RAW-Formate nur sehr eingeschränkt zur Archivierung empfohlen werden. Es kann jedoch aufgrund der Qualität erforderlich und sinnvoll sein, die RAW-Dateien neben einem für die (Langzeit-) Archivierung tauglicheren Format ebenfalls zu bewahren, zumindest solange eine Verwendung der speziellen Formate noch gegeben ist.

### *TIFF-Format*

Für die Speicherung der fertigen Bilddateien gilt es eine Abwägung der sich teilweise widersprechenden Eigenschaften Speicherplatzbedarf <=> Qualität <=> Kompatibilität zu treffen. Dies wird v.a. beim TIFF-Format mit seinen Optionen und Varianten deutlich.

TIFF hat sich insbesondere im Bereich der Druckindustrie als Quasistandard für Rasterbilder durchgesetzt. Es enthält allerdings außerordentlich viele Optionen und Varianten (Farbtiefe, Komprimierung, Bit-Anordnung, Ebenenkomprimierung, Farbkanäle, ...) und dies führt sehr schnell zu Inkompatibilitäten. Ein Ausweg ist die Mindestspezifikation „Baseline-TIFF“ dar, die alle „TIFF-fähigen“ Programme umsetzen sollten. Leider bleibt bei der Nutzung von Bildverarbeitungsprogrammen zumeist im Dunklen, welche Speicher-Optionen „Baseline“ sind und welche nicht.

Unkomprimierte TIFFs sind vorzuziehen, weil ein Bitfehler nicht gleich zum Totalausfall führt. Wenn die Speichergröße jedoch minimiert werden muss, kann die G4-Komprimierung (für reine Schwarz-Weiß-Bilder) bzw. die LZW-Komprimierung empfohlen werden.

Tiff-Dateien bieten außerdem beste Möglichkeiten Metadaten (EXIF, IPTC und XMP) sowie ICC Profile in die Bilddatei zu integrieren.

### *JPEG-Format*

JPEG ist normalerweise ein Bildformat mit verlustbehafteter Komprimierung, es gehen also Informationen verloren. Dies ist unter strengen archivalischen Gesichtspunkten schwer hinnehmbar. Man kann sich jedoch bei den derzeit erreichbaren Bildauflösungen auf den Standpunkt stellen, dass diese Verluste in bestimmtem Maße akzeptabel sind. Dies wird vor allem dann der Fall sein, wenn die Aufnahmen eher einen visuellen Eindruck verschaffen sollen und der exakte „messtechnische“ Effekt unbedeutend ist. In solchen Fällen kann eine moderate Komprimierung viel Speicherplatz sparen.

Deutlicher Vorteil von JPEG ist der offene Standard und die hohe Verbreitung. Sämtliche Digitalkameras mit Ausnahme einiger High-End-Kameras speichern ausgearbeitete Bilddateien im JPG-Format. Ebenso verarbeiten auch fast alle Bildverarbeitungsprogramme JPEG-Daten. Nach heutigem Stand bietet das JPG-Format die größte Wahrscheinlichkeit dass diese Dateien auch in vielen Jahren ohne weitere Konvertierungen noch zu öffnen sein werden.

Einschränkend muss jedoch darauf verwiesen werden, dass durch die verlustbehaftete Komprimierung bei jedem Speichern ein neuer, zusätzlicher Fehler entstehen kann. Insofern sollte man während der Bearbeitungsphase der Bilder das JPEG-Format möglichst vermeiden und nur die fertig ausgearbeitete Bilddatei in JPEG speichern.

Das JPEG-Format kann unter denselben Gesichtspunkten auch zur Speicherung von Bildern innerhalb des PDF/A-Formats verwendet werden.

Im Gegensatz zum „normalen“ JPEG bietet JPEG2000 eine verlustfreie Komprimierung an. JPEG2000 ist ebenfalls ein offener Standard, aber bei weitem nicht so verbreitet wie JPEG und kann daher auch nicht vorbehaltlos empfohlen werden.



**Formate für Bild-Dokumente****bevorzugt:****TIFF unkomprimiert** (TIFF 6.0 Baseline-Formate)**TIFF G4-komprimiert** (Gruppe 4, verlustfrei nach ITU/CCITT), nur für Schwarz-Weiß-Bilder direkt erzeugt (ohne z.B. JPEG-Zwischenschritt)**akzeptiert (ggf. problematisch):**ggf. **TIFF LZW**-komprimiertggf. **JPEG** (wenn Komprimierungsverluste akzeptabel sind, hohe Auflösung erforderlich)**Kandidaten (problematisch, weiter prüfen):**ggf. **JPEG2000** (verlustfreie Komprimierung) (offen aber noch nicht verbreitet, keine Erfahrungen)ggf. **DNG** als Rohdatenformat (standardisiert aber nicht weit verbreitet, keine Erfahrungen)

Zusammenfassung: Bild-Formate zur Archivierung

**7.3. CAD-Daten**

(David Bibby, Reiner Göldner)

Sowohl im CAD-Bereich als auch bei GIS gibt es eine große Heterogenität der Datenformate und nur wenige Standards, die jedoch nicht weit verbreitet sind. In dieser Situation findet man keine Formate, die unproblematisch für die Archivierung geeignet sind, so dass auch keine Empfehlung für ein „bevorzugtes“ Archivierungsformat ausgesprochen werden kann. Nachfolgend sollen daher die verbreitetsten Formate betrachtet und hinsichtlich ihrer Archivierung bewertet werden.

Obwohl es unter Anwendung gängiger Programme (ArchäoCAD, TachyCAD, Singularch) mehr oder minder erfolgreich möglich ist, Sachdaten an CAD-Geometrien anzuhängen, ist diese Arbeitsweise nicht sehr üblich (und wird zumeist GIS-Systemen überlassen). In diesem Sinne sollen CAD-Dateien als Grabungszeichnungen von Flächen und Profilen, übergreifenden Grabungszusammenhänge usw. verstanden werden, die keine zusätzlichen Sachdaten enthalten.

Hier wird auf die Archivierung und Gewährleistung der Leserlichkeit und inhaltliche Integrität der Daten eingegangen – die Problematik des Austausches mit anderen Systemen (vor allem mit GIS) ist hinlänglich bekannt. Diese Probleme sind aber nicht unbedingt die gleichen, die bei der Frage der Archivierung von CAD-Daten auftauchen. Wenn es um Archivierung geht, ist die Situation sowohl einfach wie auch problematisch.

Weitere Informationen findet man z.B. im Aufsatz „Preserving Computer-Aided Design“ (→ Alex Ball).

**AutoCAD-DWG:**

Der CAD Marktführer ist mit Abstand Autodesk mit dem unter Windows laufenden Produkt AutoCAD. In Folge dessen ist das „DWG“ (Drawing) Format von AutoDESK zum quasi-Standard für 2D- und 3D-CAD-Zeichnungen geworden. Dieses Format ist völlig proprietär, es ändert sich mit jeder Programmversion (jährlich) und die Dokumentation ist nicht frei erhältlich. Neue Programmversionen sind abwärtskompatibel – zumindest einige Stufen. Um die vollständige Funktionalität von AutoCAD zu gewährleisten, ist man zumeist auf das DWG-Format angewiesen.

**AutoCAD-DXF:**

Das DXF-Format ist ebenfalls ein proprietäres Format von Autodesk, allerdings ist die Dokumentation für DXF frei verfügbar, was zu einer weiten Verbreitung und zu großer Akzeptanz als Zeichnungsaustauschformat geführt hat. Damit ist DXF ein Quasistandard. Leider ändert sich auch das DXF-Format mit jeder neuen Programmversion von AutoCAD.

DXF ist ein in ASCII-Zeichen lesbares Abbild der binär abgespeicherten DWG und unterstützt direkt 2D und 3D Koordinaten. Der Aufbau ist sehr klar, einfach und strukturiert. DXF ist oft der kleinste gemeinsame Nenner vieler Vektorgrafikprogramme, aber meist werden damit nicht alle Funktionen dieser Programme voll unterstützt und es können beim Export Daten oder Funktionalitäten verloren gehen.

Aufgrund des offenen Standards und der weiten Verbreitung ist DXF zur Archivierung recht gut geeignet, mit der Einschränkung der geringen Haltbarkeit und der damit verbundenen Notwendigkeit häufi-



ger Migrationen. Das DXF-Format R12 (1992) ist beispielsweise mit vielen CAD- und GIS-Systemen lesbar und damit für die Archivierung geeignet.

### Formate für CAD-Dokumente

**bevorzugt:**

derzeit keine Empfehlung möglich

**akzeptiert (ggf. problematisch):**

ggf. **DXF** (Einschränkungen gegenüber der aktuellen Funktionalität möglich, DXF R12 z.B. ist ein gut lesbares Format)

**Kandidaten (problematisch, weiter prüfen):**

ggf. **DWG** (stark eingeschränkte Nutzungsdauer, häufige Migrationen erforderlich)

ggf. **PDF/A** (nur für Ansichten/Visualisierungen mit stark beschränkter Funktionalität)

Zusammenfassung: CAD-Formate zur Archivierung

Alle weiteren CAD-Formate spielen eine weit untergeordnete Rolle und kommen daher für die Archivierung nicht in Frage. Daher sollten CAD-Daten möglichst als DWG und DXF gespeichert werden. Etwaige CAD-fremde externe Referenzen (Bitmaps etc.) unterliegen den Empfehlungen für die jeweilige Datengattung.

Das Grundproblem bei der Archivierung von CAD-Daten ist der schnelle Wandel der Versionen, der eine häufige Migration erfordert. Wenn man sich auf längere Migrationszyklen einlässt (und dabei ggf. mehrere Versionen überspringt), geht dies mit einem Funktionsverlust gegenüber den aktuellen Programmversionen einher. Am sinnvollsten kann man dem begegnen, indem man mehrere Versionen für verschiedene Nutzungszwecke parallel vorhält:

- Eine möglichst stabile und weit verbreitete (ggf. ältere) Version (Vorschlag DXF) wird zur dauerhaften Archivierung (laufende Migration) benutzt, wobei der wesentliche Funktionsgehalt erhalten bleiben sollte, jedoch spezielle Eigenschaften verloren gehen können.
- Die bei der Erstellung aktuelle Version (Vorschlag: DWG) wird begrenzt aufbewahrt, um eine verlustfreie Weiterbearbeitung zu ermöglichen, solange noch Lesbarkeit besteht (dies ist in unserem Sinn eigentlich keine Archivierung).
- Eine weitgehend problemlose Archivierung lässt sich durch die Erstellung von Ansichten (Visualisierungen, Ausdrucke) über Text- oder Bildformate erreichen (vgl. Abschnitte 7.1 und 7.2., Empfehlung PDF/A), möglicherweise ist auch die Erstellung von Hardcopys (vgl. Abschnitt 6.1) sinnvoll. Hierbei ist auf eine sorgfältige Einstellung der möglichen Ansichten zu achten, damit möglichst viele Aspekte der Originaldaten wiedergegeben werden (ggf. auch an Listen von Attributen und Sachdaten denken)! Die originale Funktionalität der Daten geht dabei jedoch verloren!

Mit dieser Methode wird sowohl die dauerhafte Archivierung (ggf. mit inhaltlichen Abstrichen), als auch die vollinhaltliche Nutzbarkeit (zumindest für einen begrenzten Zeitraum) gewährleistet.

CAD-Daten werden umfangreicher Pflege bedürfen, sie unterliegen einem hohen Migrationsaufwand. Diese Migration kann automatisch durchlaufen, muss aber zumindest stichprobenartig manuell auf Datenverluste kontrolliert werden. Dazu ist es erforderlich, dass die CAD-Daten möglichst schnell eine geeignete „Archivierungs- und Pflegeeinrichtung“ erreichen, wo sie entsprechend behandelt werden. Dies könnte eine entsprechend ausgerüstete behördeninterne Einrichtung oder auch eine kompetente externe Stelle sein.

## 7.4. Geodaten

(Axel Posluschny, Reiner Göldner)

Im GIS-Bereich unterscheidet man zunächst einmal zwischen Vektordaten und Rasterdaten, wobei die Vektordaten am ehesten GIS-typisch sind und die Rasterdaten oft aus anderen Quellen mit deren typischen Formaten stammen (z.B. Bildformate). Die Verwendung dieser Daten in einem GIS ist höchst komplex und auf die jeweilige Funktionalität direkt bezogen. Eine Standardisierung, wie sie für die Archivierung wünschenswert ist, setzt dieser Funktionalität klare Grenzen und konnte sich daher



bisher nur unzureichend durchsetzen. Es gibt jedoch einige Datenformate, die sich als eine Art Quasi-Standard etabliert haben und von vielen GIS importiert und genutzt werden können.

Aufgrund der hohen Komplexität und geringen Standardisierung kann derzeit keine uneingeschränkte Empfehlung für die Archivierung von GIS-Daten ausgesprochen werden. Demzufolge erfordert die Archivierung eine hohe Aufmerksamkeit und einen hohen Migrationsaufwand, ähnlich wie bei CAD-Daten

#### *AutoCAD DXF:*

Die meisten GIS unterstützen den Import und Export von CAD-Daten, wobei dem DXF-Format dabei eine herausragende Stellung zukommt. Damit können auch GIS-Daten als DXF archiviert werden, mit allen schon benannten Problemen (s. Abschnitt 7.3). Ein Nachteil ist die Tatsache, dass DXF aus sich heraus keine Sachinformationen übermitteln kann, wie sie üblicherweise an GIS-Zeichnungsobjekte angehängt sind. Zusätzliche Informationen, wie z.B. die Beschreibung des Befundes, sind meist nicht ohne Weiteres mit dem Befundzeichnungsobjekt in einem CAD-System verknüpfbar (dies ist meist teurer Spezialsoftware vorbehalten). Aus diesen Gründen ist eine Speicherung von GIS-Vektordaten im DXF-Format wenig empfehlenswert. Die Sachinformationen müssen in jedem Fall zusätzlich archiviert werden, z.B. in einem Tabellenformat.

#### *Tabellenformate:*

Zumindest die Sachinformationen eines GIS lassen sich zumeist über gängige Datenbankformate importieren und exportieren (TXT, CSV, XLS, ..., gelegentlich auch DBF, MDB, XML u.a.). In Spezialfällen funktioniert dies auch mit den Geometrie-Informationen, z.B. für Punktkoordinaten oder bestimmte textbasierte Repräsentationen der Geometrie (Simple Feature Access bzw. Well Known Text - WKT). In diesen begrenzten Fällen kann eine Archivierung nach den Kriterien von Datenbanken erfolgen (vgl. Abschnitt 7.5). Dabei ist aber darauf zu achten, dass die Punktinformation in einer aus einem GIS zu exportierenden Datei nicht automatisch die beiden benötigten Spalten für das Koordinatenpaar enthält, diese müssen nötigenfalls vorher erst erzeugt und mit den entsprechenden Informationen gefüllt werden.

#### *ADeX:*

Das Archäologische Datenexportformat ADeX ist in seiner auf Punktkoordinaten beschränkten Version 1.2 uneingeschränkt archivierungstauglich. Bei der Version 2, die auch Linien und Flächen unterstützt, sind zusätzlich die Hinweise zu den Formaten SHP und MIF zu beachten.

#### *ESRI-Shape-Format – SHP:*

Das gängigste Vektorformat, das heute in GIS-Programmen Verwendung findet ist das von der Firma ESRI für seine ArcGIS/ArcINFO-Produktschiene entwickelte Shapeformat (SHP). Nahezu alle gängigen GIS-Programme können SHP direkt lesen oder aber zumindest importieren. Auch zahlreiche OpenSource-GIS-Programme (u.a. gvSIG oder Quantum) verwenden SHP trotz seiner Eigenschaft als proprietäres Format. SHP ist relativ gut dokumentiert, hat sich in den letzten Jahren nicht wesentlich geändert und dadurch eine weite Verbreitung gefunden. Eine besondere Eigenschaft von Shapes ist die Tatsache, dass in einer SHP-Datei nur entweder Punkte, Multipunkte, Linien oder Flächen enthalten sein können. Eine Gewässerdatei mit Quellen (Punkten), Flüssen (Linien) und Seen (Flächen) wird also beim Export in SHP in drei einzelne Dateien aufgespalten. Zudem ist SHP nur „maschinenlesbar“, eine Ansicht mittels eines Texteditors fördert (anders als z.B. bei WKT, GML oder MIF) keine Informationen zu Tage, die auch nach einem System- bzw. Formatwechsel noch verwertbar sind. Shapefiles bestehen immer aus mehreren Dateiteilen, die Geometriedaten, Indexdaten, Sachdaten (im DBF-Format, das kann je nach Zeichensatz Probleme mit Umlauten und Sonderzeichen mit sich bringen) sowie weitere Informationen enthalten.

Trotz aller Nachteile sprechen die weite Verbreitung, die große Akzeptanz und die relativ stabile Versionierung für SHP als Archivformat, sofern auch hier die Grundsätze einer Formatmigration nötigenfalls berücksichtigt werden.



### *MapInfo Interchange Format – MIF:*

Die Software MapInfo verwendet zum Austausch mit anderen GIS-Programmen das Austauschformat MIF (MapInfo Interchange Format). Es handelt sich um ein reines Textformat, in dem (geo)graphische Informationen abgespeichert werden. Dabei werden neben formatbeschreibenden Informationen alle in einer Datei enthaltenen Zeichnungsobjekte (Punkte, Linien, Flächen) beschrieben und mit Koordinaten sowie dem verwendeten Projektionssystem abgespeichert. Aus der Textdatei sind also die relevanten Informationen auch ohne ein kompatibles GIS-Programm entnehmbar und gegebenenfalls anderweitig verwendbar. Ergänzt werden die eigentlichen MIF-Dateien durch gleichnamige MID-Dateien, die die Attribute (d.h. die Datenbankinformationen) der (geo)graphischen Objekte enthalten. Die Verbreitung als Im-/Exportformat für die in der Archäologie relevanten GIS-Produkte ist mittelmäßig.

Beim Export aus anderen Programmen als MapInfo ist darauf zu achten, dass z.B. IDRISI nicht immer die richtige Georeferenzinformation in die Datei speichert (manchmal Nicht-Welt-Meter statt GK3), diese kann aber in einem Texteditor (oder bei mehreren Dateien durch die Verwendung des Free-ware-Programms BTRWIZ) ersetzt werden.

### *XML /GML /KML:*

Außer GoogleEarth, das nicht als GIS sondern als GIS-Dataviewer mit beschränktem Formatrepertoire bezeichnet werden muss, beherrschen bislang nur wenige gängige GIS-Programme den direkten Im- und Export von XML oder seinen Derivaten GML und KML (ESRI benutzt z.B. XML zur Erzeugung bzw. Speicherung von Metadaten, MapInfo kann über ein Zusatzmodul Daten als KML exportieren). Derzeit würde ich keine Speicherung in diesen Formaten empfehlen, da ein Austausch nicht immer problemlos gewährleistet ist. Da aber die Formate gut dokumentiert und als Textdateien einfach lesbar sind und zukünftig mit einer breiteren Akzeptanz dieser Formate auch bei Softwareherstellern und -programmierern zu rechnen ist, ist eine Archivierung z.B. als GML durchaus sinnvoll, sofern das verwendete GIS-Programm einen problemlosen Export gewährleistet.

### *Sonstige ( IDRISI , GRASS , ...)*

Zahlreiche GIS-Programme sowohl aus dem kommerziellen, wie auch aus dem OpenSource-Bereich verwenden jeweils eigene Dateiformate, die mehr oder weniger verbreitet und kompatibel sind. Meist bieten diese Programme aber auch die Möglichkeit des Exports in einem der o.g. Formate. Diese Formate sollten in jedem Fall – zumindest immer zusätzlich – für die Archivierung von GIS-Vektordaten verwendet werden. Nur sie gewährleisten einen einigermaßen reibungslosen Austausch mit anderen Programmen bzw. die langfristig lesbare Speicherung und Archivierung der Informationen.

### *GPS-Daten*

Auch GPS-Daten sind letztlich meist Vektordaten (Tracks, Routes, Waypoints). Sie liegen oft in Formaten vor, die die jeweilige Hardware und die weiterverarbeitende Software vorgibt. Als einigermaßen kompatibel hat sich GPX (GPS Exchange Format) erwiesen, das auch auf einem XML-Schema aufbaut und von zahlreichen Geräten und Softwareprodukten gelesen und erzeugt werden kann. Glücklicherweise sind zahlreiche Konvertierungsprogramme für die unterschiedlichen GPS-Formate auch frei erhältlich, die umfangreichste Auswahl an unterstützten Formaten bietet dabei GPSBabel.

Selbstverständlich ist auch hier eine Speicherung als GML oder KML oder eine Umwandlung in eines der anderen oben empfohlenen GIS-Formate (SHP, MIF) möglich und sinnvoll.

### *Geländemodelle*

Auch Geländemodelle können außer als Raster- auch als Vektordaten gespeichert werden. Zum Teil sind die Übergänge hier fließend. Einige der am weitesten verbreiteten Formate sind Tabellen, die die entsprechenden Höheninformationen enthalten.

Unregelmäßige Informationen können z.B. als einfache Tabelle mit den Spalten X-Koordinate, Y-Koordinate und Höhe vorliegen und später gegebenenfalls im GIS in regelmäßige Raster umgesetzt (interpoliert) werden. In dieser Form lassen sich auch regelmäßig verteilte Werte (regelmäßige Messraster und vor allem interpolierte Daten) abspeichern. Aus Platzgründen und weil entsprechende Dateien von fast allen gängigen GIS-Programmen gelesen und erzeugt werden können empfiehlt sich



aber die Verwendung des sog. nicht proprietären ESRI- bzw. ARC/INFO-ASCII-Grids. Dabei werden nach einem standardisierten Header, der u.a. Informationen zur Zeilen- und Spaltenzahl des Rasters und zur Zellgröße sowie den Eckkoordinaten enthält, die Höhenwerte (oder sonstigen Rasterinformationen) zeilenweise in der Datei fortgeschrieben).

Da diese Daten als in einem Texteditor anzeigbare und veränderbare Dateien vorliegen, empfiehlt sich neben der Speicherung bzw. Archivierung in einem GIS-spezifischen Rasterformat (GeoTIFF, DTED, ...) auch einen Export in ein einfaches Tabellenformat, wie oben beschrieben.

#### *GeoTIFF u.a. Georeferenzierte Rasterdaten*

Rasterdaten benötigen zur Darstellung im GIS eine Georeferenz, d.h. Koordinaten für die Lage und Ausdehnung sowie das benutzte Koordinatenreferenzsystem. Neben verschiedenen proprietären Formaten ist GeoTIFF weit verbreitet. GeoTIFF ist eine spezielle Form des TIFF-Formats. Die normalen TIFF-Optionen bleiben erhalten (vgl. Abschnitt „Bilder“), so dass die Dateien theoretisch auch mit normalen Viewern lesbar sind (allerdings werden oft nicht die Baseline-Optionen benutzt, so dass doch wieder spezielle Viewer erforderlich werden). Zusätzlich werden die Georeferenz-Informationen in speziellen Tags abgelegt, die mittlerweile von den meisten GIS-Systemen ausgewertet werden können.

Aus der ESRI-Welt (ArcGIS) stammt das Verfahren, eine normale Bilddatei um ein sog. „World File“ zu ergänzen, welches die Georeferenz in lesbarer Form (TXT) enthält, allerdings ohne Koordinatenreferenzsystem. Letzteres muss also separat übermittelt und bewahrt werden. World-Files werden typischerweise für JPEG-, PNG-, GIF- und TIFF-Bilddateien benutzt.

Beide Varianten sind offene Quasi-Standards und weit verbreitet, aber trotzdem nicht immer unproblematisch. Falls angeboten, sollte man ein GeoTIFF im Baseline-Format benutzen und zusätzlich noch ein World-File erstellen, das benutzte Koordinatenreferenzsystem gehört dann in die Metadaten.

#### *Amtliche Landesvermessung*

Die Vermessungsämter der deutschen Bundesländer erzeugen Daten in eigenen Formaten bzw. als eigene Modelle (z.B. ALK, AFIS-ALKIS-ATKIS). Diese werden über die sogenannte Einheitliche Datenbankschnittstelle (EDBS) ausgetauscht, die jedoch in gängigen GIS-Programmen bislang keine Rolle spielt.

#### **Formate für GIS**

##### **bevorzugt:**

derzeit keine Empfehlung möglich

##### **akzeptiert (ggf. problematisch):**

ggf. **SHP** (weit verbreitetes proprietäres, aber offen beschriebenes Binärformat)

ggf. **ADeX** (für Punktinformationen bzw. in Verbindung mit SHP oder MIF)

##### **Kandidaten (problematisch, weiter prüfen):**

ggf. MIF (proprietäres, aber offen beschriebenes Textformat)

ggf. GML (standardisiertes XML-basiertes Format, wenig verbreitet)

ggf. KML (proprietäres, offen beschriebenes XML-basiertes Format)

ggf. PDF/A (nur für Ansichten/Visualisierungen mit stark beschränkter Funktionalität)

ggf. GeoTIFF (besser mit zusätzlichem World-File und Angabe zum Koordinatenreferenzsystem)

Zusammenfassung: GIS-Formate zur Archivierung

Das Grundproblem bei der Archivierung von GIS-Daten ist, wie schon bei CAD-Daten erwähnt, der schnelle Wandel der Versionen, der eine häufige Migration erfordert. Wenn man sich auf längere Migrationszyklen einlässt (und dabei ggf. mehrere Versionen überspringt), geht dies mit einem Funktionsverlust gegenüber den aktuellen Programmversionen einher. Die Speicherung mehrerer Versionen in unterschiedlichen Formaten ist auch im GIS-Bereich sinnvoll:

- Ein möglichst stabiles und weit verbreitetes Format (Vorschlag SHP) wird zur dauerhaften Archivierung (laufende Migration) benutzt, wobei der wesentliche Funktionsgehalt erhalten bleibt.
- Die bei Beginn der Archivierung für den Betrieb maßgeblichen Dateien (aktuelle Version) werden im Original begrenzt aufbewahrt, um eine verlustfreie Weiterbearbeitung zu ermöglichen, solange noch Lesbarkeit besteht (dies ist in unserem Sinn eigentlich keine Archivierung).



- Eine weitgehend problemlose Archivierung lässt sich durch die Erstellung von Ansichten (Visualisierungen, Ausdrucke) über Text- oder Bildformate erreichen (vgl. Abschnitte 7.1 und 7.2., Empfehlung PDF/A), möglicherweise ist auch die Erstellung von Hardcopies (vgl. Abschnitt 6.1) sinnvoll. Hierbei ist auf eine sorgfältige Einstellung der möglichen Ansichten zu achten, damit möglichst viele Aspekte der Originaldaten wiedergegeben werden (ggf. auch an Listen von Attributen und Sachdaten denken)! Die originale Funktionalität der Daten geht dabei jedoch verloren!

Mit dieser Methode wird sowohl die dauerhafte Archivierung (ggf. mit inhaltlichen Abstrichen), als auch die vollinhaltliche Nutzbarkeit (zumindest für einen begrenzten Zeitraum) gewährleistet.

Auch GIS-Daten werden umfangreicher Pflege bedürfen, da sie einem hohen Migrationsaufwand unterliegen und stichprobenartig manuell auf Datenverluste kontrolliert werden. Dazu ist, ebenso wie bei CAD-Daten, eine leistungsfähige und kompetente „Archivierungs- und Pflegeeinrichtung“ erforderlich.

## 7.5. Datenbanken

(Reiner Göldner)

Datenbanken sind weit verbreitete, lange und gut eingeführte IT-Systeme. Vor allem Server-Datenbanken sind gut auf einen dauerhaften Betrieb über Jahre hinweg ausgerichtet. Sie sind stark standardisiert und beinhalten die für Versionswechsel erforderlichen Migrationsmethoden. Datenbanken besitzen damit schon jetzt hohes Potential für die Langzeitarchivierung.

Aktive Datenbanken mit „lebendigen“ Daten sind dabei am problemlosesten, weil Betreuung, Wartung und auch Migration zum laufenden Betrieb gehören und damit in der Regel eingeplant und abgesichert sind. Bei inaktiven, „abgelegten“ Datenbanken sind jedoch auch Überlegungen zur Archivierung erforderlich.

Bei der Archivierung von Datenbanken ist es wichtig, dass nicht nur die Daten, sondern auch Datenmodelle und Datenbankschemata archiviert werden:

- **Datenmodell:** Die verbreiteten relationalen Datenbanken benutzen als konzeptuelles Datenmodell meist das ER-Modell (Entity Relationship Model): Entitäten stehen für Objekte der Wirklichkeit, Relationships stehen für Beziehungen zwischen den Objekten. ER-Modelle sind ausgezeichnete Modellierungsgrundlage und damit gut zur (langfristigen) Vermittlung des einer Datenbank zugrundeliegenden konzeptuellen Modells geeignet. Sie werden als Diagramm dargestellt (ERD), um die Struktur zu verdeutlichen. Eine rein textbasierte Darstellung der ERD ist nicht bekannt, daher kommt wohl am ehesten PDF/A für deren Archivierung in Frage.  
**Datenbankschema:** Das konzeptuelle Datenmodell wird als Datenbankschema technisch umgesetzt. Das Datenbankschema beinhaltet Beschreibungen der Tabellen (z.B. Name), der Attribute (z.B. Datentypen und Länge, Indizes) und Integritätsbedingungen (Primär-/Sekundärschlüssel, Referenzen, Existenz-/Wertebereiche) einer Datenbank. Die Darstellung erfolgt am besten per SQL (DDL, s.u.).
- **Daten:** Gut archivierungstauglich sind die „einfachen“ Datenformate wie Texte und Zahlen (die als Text darstellbar und damit speicherbar sind). Dagegen sind binäre Formate (BLOBs) ungeeignet, denn sie besitzen keine datenbankseitig standardisierte innere Struktur und können nicht als Text exportiert werden. BLOBs erfordern immer einen expliziten Exportmechanismus in extern lesbare Formate. In Datenbanken sollten also möglichst keine binär kodierten Dokumente oder Bilder integriert werden, sondern nur deren Links zu separat abgelegten Dateien in archivtauglichen Formaten. Daten können extern z.B. als SQL (DML) oder als CSV dargestellt werden.

SQL (Structured Query Language) ist eine standardisierte Sprache zur Definition, Manipulation, Abfrage und Steuerung von relationalen Datenbanken, die schon seit 1979 existiert. Man unterscheidet DDL (Data Definition Language), DML (Data Manipulation Language) und DCL (Data Control Language). SQL ist ein standardisiertes und weit verbreitetes Textformat, es eignet sich daher gut für den Datenaustausch bzw. die Migration in heterogenen Umgebungen. SQL ist für die Archivierung von Datenbankschema und Daten gut geeignet.

CSV (Comma Separated Value) ist ein textbasiertes Datenformat zur einfachen Speicherung von Tabelleninhalten. Es ist zwar nicht standardisiert, aber aufgrund seiner großen Verbreitung und Akzeptanz als Quasi-Standard zu betrachten. Eine Beschreibung findet man in RFC 4180 (Request for



Comments – techn./org. Dokumente zum Internet). CSV ist für die Archivierung von Tabellen-Daten gut geeignet.

Desktop-Datenbanken wie z.B. MS-Access sind nicht zur direkten Archivierung geeignet. Ihre internen Strukturen können sich von Version zu Version schnell ändern. Sie müssen wie o.a. mit Datenmodell, Datenbankschema und Daten als TXT o.ä. archiviert werden oder in eine aktive Server-Datenbank mit stabiler Struktur übernommen werden.

Ggf. ist XML (mit DTD) zukünftig eine Alternative bei der Archivierung von Datenbankschemata und Daten. XML ist textbasiert und daher gut archivierungstauglich. Gegenwärtig liegen aber noch nicht viele Erfahrungen mit XML-Daten vor, die Thematik sollte also weiter verfolgt werden.

Das Schweizer Bundesarchiv hat einen eigenen Standard „SIARD RDB DATA“ etabliert. Tools extrahieren dabei aus relationalen Datenbanken Strukturinformationen und Daten und wandeln sie in ein archivierbares Format um. Es wird auch eine Beschreibung der Tabellen und Attribute erwartet. Der Standard basiert auf SQL, CSV und XML.

(Übrigens: Das Schweizer Bundesarchiv archiviert ausschließlich die folgenden Formate: TXT, PDF/A, CSV, SIARD, TIFF und WAVE.)

Das „kopal“-Projekt (Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen) vertraut in die Langzeitarchivierungsfähigkeit professioneller Datenbanksysteme (IBM). Allgemeine Vor- und Nachteile wurden ja schon beschrieben. Im Falle von „kopal“ werden binäre Dokumentformate als BLOBs gespeichert, womit sich eine Abhängigkeit von proprietären Datenbankschnittstellen ergibt.

Das Deutsche Archäologische Institut (DAI) fordert zur Archivierung von Datenbanken eine Dokumentation (Anforderungen, Datenstruktur als ERD, Tabellen, semantische Beschreibung). Es sollen keine BLOBs, sondern Referenzierungen auf das Verzeichnissystem benutzt werden. Neben der Original-Datenbank soll auch ein softwareunabhängiges Form als XML (+ dokumentierte DTD) oder SQL (DDL) bereitgestellt werden.

Weitere Hinweise zur Archivierung von Datenbanken, auch hinsichtlich XML, findet man bei „Nestor“.

### Formate für Datenbanken

**bevorzugt:**

dauerhaft betriebene Online-Datenbank

**akzeptiert (ggf. problematisch):**

ggf. **SQL** (Modell und Schema beifügen)

ggf. **CSV** (Modell und Schema beifügen)

**Kandidaten (problematisch, weiter prüfen):**

ggf. PDF/A (nur für Modell, auch für Ansichten/Visualisierungen mit stark beschränkter Funktionalität)

Zusammenfassung: Datenbank-Formate zur Archivierung

Der Betrieb einer Datenbank (Server-DB) beinhaltet in sich schon alle Aspekte der Langzeitarchivierung, vor allem die Migration. Insofern muss eine aktive, laufende Datenbank nicht archiviert werden. Bei der Archivierung von Datenbanken müssen neben den Daten unbedingt auch die Datenmodelle und Datenbankschemata archiviert werden. Geeignete Formate sind SQL und CSV für Daten und Datenbankschema sowie ggf. PDF/A für das konzeptionelle Datenmodell.

## 8. Erforderliche Ressourcen

(Reiner Göldner)

Zur Archivierung ist ein angemessener Betreuungsaufwand erforderlich. Dies sind wir von herkömmlichen Archiven schon gewohnt, wo Klimatisierung und Brandschutz sowie personelle Betreuung des Bestandes und ggf. Restaurierungsleistungen üblich und gut akzeptiert sind. Es ist irrig anzunehmen, dies sei bei digitalen Archiven anders. Im Gegenteil, die vorgestellten Methoden erfordern einen permanenten Betreuungsaufwand, da die Migrationszyklen so sehr kurz sind.

Betreuungsaufwand ist in mehrerer Hinsicht erforderlich:

- organisatorisch zur Planung, Durchführung und Kontrolle von Maßnahmen,
- technisch zur Betreuung von Hard- und Softwaresystemen sowie Datenträgern und Daten (wofür entsprechendes Know-How erforderlich ist),
- personell (qualifiziert und kompetent)
- und finanziell.

*Zur Ergänzung dieses Abschnitts sind noch die folgenden Themen von Interesse:*

- *deutlicher Hinweis: Archivierung = laufende Kosten!*
- *Abschätzung technischer, personeller und finanzieller Ressourcen*
- *Unterscheiden: Selbst machen oder als Dienstleistung einkaufen*
- *hohe Qualitätsansprüche (z.B. Bildauflösung) erfordern ggf. schon bei der Erfassung mehr Ressourcen*
- *Hinweis auf ADS (arbeitet nach wirtschaftlichen Prinzipien)*
- *Die Kosten der Archivierung müssen vorab bedacht werden und schon bei der Entstehung der Daten kalkuliert werden (= Projektkosten)!*
- *Ggf. muss ein Kosten/Nutzen-Verhältnis aufgestellt werden (Wieviel ist mir die Archivierung dieses Dokuments wert?) Dabei müssen ggf. die Interessen der Produzenten und des Archivs ausgeglichen werden.*

## 9. Archivierungserfahrungen

Bei einer kleinen Umfrage der AG Archivierung des VLA wurden folgende für die Langzeitarchivierung digitaler Daten taugliche Softwaresysteme angegeben: Digitales Archiv NRW Software suite (DiPS, DNS), ExLibris Rosetta, Archivematica, iRods, Fedora/Islandora, MyCoRe, dSPACE, Mediafiler (eher Dokumentendatenbank). Weiterhin ergab sich die folgende Liste mit potentiellen Dienstleistern für die Archivierung digitaler Daten:

- IANUS-Forschungsdatenzentrum (<http://www.ianus-fdz.de>)
- Fa. Arctron (<http://www.arctron.de>)
- ADS: Archaeology Data Service (<http://archaeologydataservice.ac.uk/>)
- EDNA: E-depot voor de Nederlandse archeologie (<http://edna.leidenuniv.nl/>)
- DA-NRW: Digitales Archiv Nordrhein-Westfalen (<https://www.danrw.de/>).

Außerdem beschäftigen sich viele große Bibliotheken oder Bibliotheksverbünde mit dem Thema Archivierung digitaler Daten, u.a. die SLUB Dresden (<http://www.slub-dresden.de/ueber-uns/slubarchiv/>). Natürlich sind auch die Archive mit diesem Thema befasst, z. B. das Bundesarchiv in Deutschland (<https://www.bundesarchiv.de/fachinformationen/00895/index.html.de>).

Diese kurze Aufzählung ist natürlich nicht vollständig, aber sie mag als Anregung für weitere eigene Recherchen dienen.

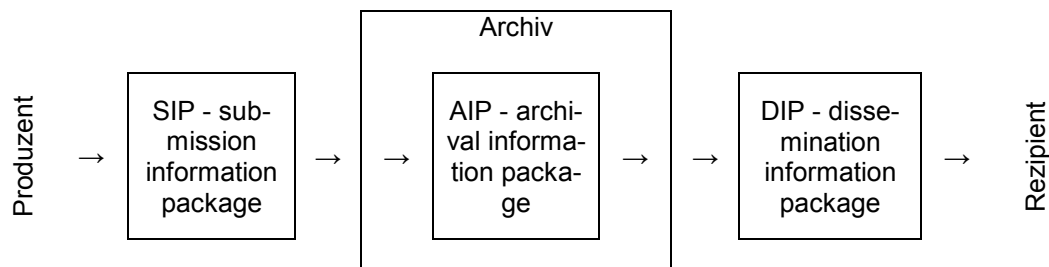


## 10. OAIS - Open Archival Information System

(Reiner Göldner)

OAIS bedeutet „Open Archival Information System“. Es ist ein Referenzmodell für ein digitales Archiv und bildet den Standard ISO 14721. Als konsequent logisches Modell ist es unabhängig von jeder Implementierung. Es leistet einen wesentlichen Beitrag zum (einheitlichen) Verständnis digitaler Archivierung. Viele Archivierungskonzepte gründen auf den Entwürfen des OAIS.

OAIS geht aus Datensicht von sog. Informationspaketen aus, die neben dem eigentlich zu bewahrenden Inhalt noch beschreibende Information (z.B. zum Datenformat) beinhalten. Die zur Archivierung eingelieferten Pakete (SIP - submission information package) werden bei der Übernahme zu Archiv-Informationspaketen (AIP - archival information package) umgewandelt. Diese beinhalten u.a. auch Informationen zur Historie der bisherigen Bearbeitungen bzw. Migrationen im Archiv sowie Attribute, die Integrität und Authentizität gewährleisten können. Auch Recherche und Zugriff erfolgen über spezielle Informationspakete (DIP - dissemination information package), die entsprechend der Anfrage und den Zugriffsrechten unterschiedlich ausgeprägt sein können. Nachfolgende Abbildung verdeutlicht dieses Modell:



**OAIS beschreibt ein digitales Archiv als Organisationseinheit von Mensch und technischen Systemen, die dazu dient, digitale Informationen dauerhaft zu erhalten und verfügbar zu machen. Es wird mittlerweile weltweit als Referenzmodell und Standard akzeptiert und angewandt.**

Funktionell definiert OAIS sechs Aufgabenbereiche, die nachfolgernd kurz beschrieben werden.

### 10.1. Datenübernahme (Ingest)

Neben der physischen Übernahme der Daten (SIPs) sind weitere Schritte erforderlich. Die Daten müssen auf Archivtauglichkeit, Vollständigkeit und Unversehrtheit geprüft werden. Es müssen sowohl die inhaltlich als auch die technisch beschreibenden Informationen extrahiert werden. Aus all diesen Informationen werden AIPs generiert und im Archiv gespeichert (→ Datenaufbewahrung). Zugleich wird die → Datenverwaltung über den Neuzugang informiert.

### 10.2. Datenaufbewahrung (Storage)

Die Datenaufbewahrung sorgt dafür, dass die AIPs physisch erhalten bleiben. Dazu dienen vielfältige Maßnahmen der IT-Sicherheit, wie redundante Speicherung, Backup und Restore sowie regelmäßige Prüfungen der Datenintegrität. Auf Anfrage werden die AIPs ausgelesen und an den → Zugriff weitergegeben.

### 10.3. Datenverwaltung (Management)

Die Datenverwaltung ist für die inhaltlich und technisch beschreibenden Informationen der AIPs zuständig. Über sie können Archivbestände identifiziert werden, sie nimmt Rechercheanfragen entgegen, bearbeitet diese und organisiert den → Zugriff. Die technischen Informationen sind für die → Erhaltungsplanung interessant und wichtig.

#### **10.4. Zugriff (Access)**

Der Bereich Zugriff unterstützt den Nutzer des Archivs bei der Recherche nach den gewünschten Informationen. Hier werden Anfragen entgegengenommen, verarbeitet und die Ergebnisse in DIPs umgewandelt und dem Rezipienten entsprechend seiner Berechtigungen zur Verfügung gestellt. Die DIPs können dabei, anders als in klassischen Archiven, in verschiedenen Ausprägungen generiert werden (z.B. als Bildschirm-Vorschau und als hochauflöste Druckversion).

#### **10.5. Erhaltungsplanung (Preservation Planning)**

Die Erhaltungsplanung bezieht sich einerseits auf den technologischen Fortschritt im Bereich der digitalen Archivierung und damit auf das Archivsystem selbst. Andererseits dient die Erhaltungsplanung der Entwicklung und Umsetzung konkreter Erhaltungsmethoden. So muss z.B. die Erneuerung veralteter AIP-Datenformate organisiert werden, wobei Fragen der Format-Konvertierung sowie des Erhalts von Integrität und ggf. Rechtsverbindlichkeit bedeutsam sind. Damit kommt diesem Bereich konzeptionell eine zentrale Bedeutung zu.

OAIS ist stark auf die Migrationsmethode zugeschnitten, bleibt jedoch auch für andere Ansätze (z.B. Emulation) offen.

#### **10.6. Systemverwaltung (Administration)**

Die Systemverwaltung betrifft das Archiv als Gesamtsystem. Es werden Konfigurationseinstellungen verwaltet, die Beziehungen der Komponenten untereinander organisiert und Zugriffsrechte überwacht.



## 11. Schlussfolgerungen, Empfehlungen

(Reiner Göldner)

**Das Ziel der digitalen Archivierung ist es, digitale Inhalte oder digitale Funktionalitäten dauerhaft zu erhalten und verfügbar zu machen.**

Dabei treten jedoch einige Probleme auf. Während analoge Medien in der Regel direkt lesbar sind, können digitale Medien nur indirekt gelesen werden. Digitale Informationen erfordern ein komplexes System aus Hardware, Software, Datenträger und Datenformat, um einem Rezipienten zugänglich gemacht zu werden. Im analogen Bereich wirken Alterungsprozesse langsam, kontinuierlich und damit vorhersehbar. Im digitalen Bereich dagegen wirken spontane und kurzfristige Alterungsprozesse, die durch komplexe Systeme mit relativ geringer Redundanz noch brisanter werden. Vorteilhaft ist jedoch die Möglichkeit, digitale Informationen ohne Verlust kopieren und damit erneuern zu können.

Zur Vorsorge gibt es mehrere Methoden. Die Hardcopy-Methode eignet sich gut für einfache Dokumente, deren Inhalt per Ausdruck vollständig wiedergegeben werden kann. Die Hardcopies werden herkömmlich archiviert. Die Methode ist für komplexe Inhalte sowie für Funktionalitäten nicht geeignet. Die Erhaltungs-Methode eignet sich für Daten, die nur in ihrer originalen Systemumgebung benutzbar sind oder wenn Authentizität im Vordergrund steht. Sie ist sehr aufwändig und außerdem für eine dauerhafte Bewahrung nicht geeignet. Die Emulations-Methode ist aufwändig, zeichnet sich aber dadurch aus, dass sie Originaldaten nutzen kann. Weiterhin bietet sie gute Möglichkeiten, Funktionalitäten zu bewahren. Bei der Migrations-Methode werden alle Systemkomponenten, also auch die Daten, stets aktuell gehalten. Dies gewährleistet einen niedrigen Betreuungsaufwand. Die Aktualisierung der Daten erfordert ggf. eine Überprüfung der Authentizität.

Noch gibt es keine fertig konfektionierten Digitalarchive zu kaufen. Jedoch sind die Prinzipien, nach denen sie arbeiten müssen, recht klar, so dass man beginnen kann, ein digitales Archiv selbst aufzubauen und zu betreiben. Wenn man berücksichtigt, dass nicht die Lebensdauer über den Erhalt entscheidet, sondern die Methodik, und wenn man akzeptiert, dass Archivierung keine passive Ablagerung ist, sondern ein aktiver Prozess, dann wird digitale Archivierung schon jetzt funktionieren.

Die Archivierung digitaler Daten ist eine komplexe Aufgabenstellung, die Kompetenzen sowohl aus der Informationstechnik als auch aus der Archivierung erfordert! Sorgfältige Planung und Vorbereitung spielen, wie schon bei der herkömmlichen Archivierung, auch im digitalen Bereich eine wichtige Rolle, ebenso angemessene finanzielle und personelle Ressourcen.

**Archivierung ist keine passive Ablagerung, sondern ein aktiver Prozess. Nicht die Lebensdauer entscheidet über den Erhalt, sondern die Methodik, Datenträger und Datenformat sind dabei sekundär. Eine klare Konzeption sowie angemessene finanzielle und personelle Ressourcen bilden die Grundlage für einen erfolgreichen Start in die Archivierung digitaler Daten.**



## 12. Quellen/Verweise

Bei der Vorbereitung dieses Ratgebers wurden Quellen genutzt, die als Startpunkt für weitere Recherchen wärmstens empfohlen werden und bei deren Autoren wir uns für viele Anregungen bedanken:

- „Nestor“ – Kompetenznetzwerk Langzeitarchivierung  
<http://www.langzeitarchivierung.de>
- „Memorandum zur Langzeitverfügbarkeit digitaler Informationen in Deutschland“  
<http://www.langzeitarchivierung.de/downloads/memo2006.pdf>
- „nestor Handbuch: Eine kleine Enzyklopädie der Langzeitarchivierung“  
 [Version 0.1, März 2007] bzw. neu [Version 1.2, Juni 2008]  
<http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf>
- Kooperativer Aufbau eines Langzeitarchivs  
 (Deutsche Nationalbibliothek, ..., IBM Deutschland, BMBF)  
<http://kopal.langzeitarchivierung.de>
- Bundesamt für Sicherheit i. d. Informationstechnik  
 (IT-Grundschutz)  
[https://www.bsi.bund.de/DE/Themen/ITGrundschutz/itgrundschutz\\_node.html](https://www.bsi.bund.de/DE/Themen/ITGrundschutz/itgrundschutz_node.html)  
 (IT-Grundschutz-Kataloge, Baustein B 1.12 Archivierung)  
<http://www.bsi.de/gshb/deutsch/baust/b01012.htm>
- Vereinigung deutscher Wirtschaftsarchivare e.V., Arbeitskreis "Elektronische Archivierung"  
<http://www.wirtschaftsarchive.de/akea/akea.htm>
- PDF/A Competence Center  
<http://www.pdfa.org>
- „Open Archival Information System“ (ISO 14721)  
 Referenzmodell für digitales Langzeitarchiv  
<http://public.ccsds.org/publications/archive/650x0b1.pdf>
- Wikipedia  
<http://de.wikipedia.org>
- Media de Lux – Laserbelichtung auf Farbmikrofilm  
<http://www.mediadelux.de/>
- JHOVE – JSTOR/Harvard Object Validation Environment  
<http://hul.harvard.edu/jhove/>
- PRONOM – Online Informationssystem zu Dateiformaten  
<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>
- DIN – Deutsches Institut für Normung  
[www.din.de](http://www.din.de)
- BSI – Bundesamt für Sicherheit in der Informationstechnik  
[www.bsi.bund.de](http://www.bsi.bund.de)
- DAI – IT-Leitfaden des Deutschen Archäologischen Instituts  
<http://www.dainst.org/de/project/it-leitfaden>
- PREMIS - PREMIS Data Dictionary for Preservation Metadata  
<http://www.loc.gov/standards/premis/>
- ADeX – Archäologischer Datenaustausch-Standard  
[http://www.landesarchaeologen.de/fileadmin/Dokumente/Dokumente\\_Kommissionen/Dokumente\\_Archaeologie-Informationssysteme/Dokumente\\_AIS\\_ADeX/ADeX\\_2-0\\_Doku.pdf](http://www.landesarchaeologen.de/fileadmin/Dokumente/Dokumente_Kommissionen/Dokumente_Archaeologie-Informationssysteme/Dokumente_AIS_ADeX/ADeX_2-0_Doku.pdf)
- Bibby/Göldner - Archäologische digitale Daten – Authentizität und Funktionalität.  
 In: Archäologie und Informationssysteme. Vom Umgang mit archäologischen Fachdaten in Denkmalpflege und Forschung. Arbeitshefte zur Denkmalpflege in Niedersachsen 42. Niedersächsisches Landesamt für Denkmalpflege (Hameln 2013), 70-75
- Alex Ball. Preserving Computer-Aided Design (CAD)  
 DPC Technology Watch Report 13-02 April 2013:  
[http://www.dpconline.org/component/docman/doc\\_download/896-dpctw13-02pdf](http://www.dpconline.org/component/docman/doc_download/896-dpctw13-02pdf)

Abbildungen:

Trittsteine (Frontseite)

Original: „Stepping stones“ by Helen, <http://moblog.net/view/51756/stepping-stones> (22.11.2011)

Alterungsprozess (Abschnitt 1)

Landesamt für Archäologie Dresden, Diathek



**Anlage 1: Datei-Spezifikationen**

***	TXT	<ul style="list-style-type: none"> <li>für unstrukturierten und unformatierten Text</li> <li>Kodierung/Zeichensatz: <ul style="list-style-type: none"> <li>US-ASCII (keine dt. Sonderzeichen!)</li> <li>Unicode UTF-8 (oder UTF-16)</li> <li>ISO 8859-1(Latin-1 westeuropäisch)</li> </ul> </li> <li>bei strukturierten Texten → CSV, XML</li> <li>bei formatierten Texten → PDF-A, TIFF</li> </ul>
***	PDF/A	<ul style="list-style-type: none"> <li>nach ISO 19005-1</li> <li>min. Konformitätsgrad 1b (weniger streng), 1a empfohlen</li> <li>mgl. Adobe-PlugIn verwenden oder Format validieren</li> </ul>
*	CSV	<ul style="list-style-type: none"> <li>nach RFC 4180</li> <li>Kodierung wie TXT</li> <li>Spezifikation (default): <ul style="list-style-type: none"> <li>Kopfzeile verwenden</li> <li>nur ein Datensatz je Zeile</li> <li>Trennzeichen , ; oder  </li> </ul> </li> <li>abweichende Spezifikation explizit angeben (z.B. in Datei schema.ini)</li> <li>ggf. inhaltliche Beschreibung erforderlich</li> <li>ggf. Strukturbeschreibung erforderlich (ER-Modell)</li> </ul>
?	XML	<ul style="list-style-type: none"> <li>vgl. Extensible Markup Language (XML) 1.0 (Fifth Edition), W3C Recommendation 26 November 2008 (<a href="http://www.w3.org/TR/2008/REC-xml-20081126/">http://www.w3.org/TR/2008/REC-xml-20081126/</a>)</li> <li>zugrundeliegendes Textformat wie TXT</li> <li>ggf. inhaltliche Beschreibung erforderlich</li> <li>ggf. Strukturbeschreibung erforderlich (ER-Modell)</li> <li><b>Anmerkung:</b> XML-Formate beziehen sich immer auf eine spezifische Anwendung, die zur adäquaten Nutzung der Daten mit zur Verfügung stehen muss. Das Lesen der Textinfo ist eigentlich sekundär, aber ggf. hilfreich, wenn die Originalanwendung nicht zur Verfügung steht.</li> <li>ggf. spezielle Sprache separat betrachten, z.B. XML für Text + Layout → ODF ISO/IEC 26300 (Open Office)</li> </ul> <p>→ <b>vorerst nicht nutzen</b></p>
***	TIFF	<ul style="list-style-type: none"> <li>Version 6, Baseline Format</li> <li>proprietäre Erweiterungen werden ignoriert (Adobe-Enhancements u.a.)</li> <li>Komprimierung <ul style="list-style-type: none"> <li>unkomprimiert</li> <li>SW-Bilder auch G4 (Gruppe 4, verlustfrei nach ITU/CCITT)</li> </ul> </li> <li><u>keine</u> Mehrseiten-Bilder (Multipage-TIFF), <u>keine</u> Ebenen (Layer)</li> <li>24...48 Bit Farbe, 8...16 Bit SW</li> <li>Auflösung entspr. Inhalt (s.u.)</li> </ul>
*	TIFF-LZW	<ul style="list-style-type: none"> <li>wie TIFF, aber mit LZW-Komprimierung</li> </ul>
?	JPEG2000	<ul style="list-style-type: none"> <li>ISO 15444 zertifiziert</li> <li>speziell: verlustfreie Komprimierung benutzen!</li> <li>noch nicht gut verbreitet (PlugIn für Photoshop, med. DICOM, 2nd-Life, Reisepass)</li> </ul> <p>→ <b>vorerst nicht nutzen</b></p>
?	DNG	<ul style="list-style-type: none"> <li>offene Spezifikation von Adobe (<a href="http://www.adobe.com/de/products/dng/">http://www.adobe.com/de/products/dng/</a>) zur vom Hersteller unabhängigen Speicherung von RAW-Bilddaten</li> <li>noch nicht gut verbreitet (Photoshop, Silverfast)</li> </ul> <p>→ <b>vorerst nicht nutzen</b></p>
*	JPEG (JFIF)	<ul style="list-style-type: none"> <li>ISO 10918-1 bzw. CCITT Recommendation T.81 für Rasterbilder (→ Dateiformat ist eigentlich JFIF: JPEG File Interchange Format)</li> <li>beinhaltet verlustbehaftete Kompression</li> <li>nur ein Farbmodell zugelassen</li> </ul>

?	DWG	<ul style="list-style-type: none"> <li>• firmeninterne Spezifikation für AutoCAD</li> <li>• stark eingeschränkte Kompatibilität, aber weite Verbreitung</li> </ul> → <b>vorerst nicht nutzen</b>
*	DXF	<ul style="list-style-type: none"> <li>• quelloffene Schnittstelle der Fa. Autodesk (nicht neutral genormt)</li> <li>• eingeschränkte Kompatibilität, aber weite Verbreitung (zumeist 3...4 Softwaregenerationen lesbar)</li> <li>• verwendet ASCII-Text-Format (textlesbar)</li> <li>• oft (v.a. von Fremdherstellern) wird nicht die volle AutoCAD-Funktionalität unterstützt</li> </ul>
*	ADeX	<ul style="list-style-type: none"> <li>• Archäologischer Datenaustausch-Standard (Kommission AIS des VLA)</li> <li>• archäologisch-inhaltliche und technische Spezifikation (CSV-Dateien, textlesbar)</li> </ul>
*	SHP (Shapefile)	<ul style="list-style-type: none"> <li>• von der Fa. ESRI entwickeltes, offenes Dateiformat für Geodaten</li> <li>• Quasistandard für Desktop-GIS aufgrund des hohen Verbreitungsgrades</li> <li>• eigentlich eine Kollektion aus 3...9 Dateien</li> <li>• separate Shapefiles für Punkte, Linien, Flächen erforderlich</li> </ul>
?	MIF	<ul style="list-style-type: none"> <li>• Dateiformat der Fa. MapInfo für die gleichnamige GIS-Software</li> <li>• verwendet Textformat (textlesbar)</li> <li>• geringe Verbreitung</li> </ul> → <b>vorerst nicht nutzen</b>
?	GML	<ul style="list-style-type: none"> <li>• ISO 19136 (GML 3.2)</li> <li>• XML-Basierte Auszeichnungssprache für raumbezogene Objekte (Features)</li> <li>• geringe Verbreitung</li> </ul> → <b>vorerst nicht nutzen (ggf. neu bewerten)</b>
?	KML	<ul style="list-style-type: none"> <li>• von der Fa. Google für kartographische Zwecke benutztes Format (GML-ähnlich),</li> <li>• OGC-Standard seit 2008</li> <li>• außer bei Google Earth nicht stark verbreitet</li> </ul> → <b>vorerst nicht nutzen</b>
?	GeoTIFF	<ul style="list-style-type: none"> <li>• spezielle Form des TIFF-Formats mit integrierter Georeferenz</li> <li>• offene Spezifikation des OSGeo (aber kein ISO)</li> <li>• weit verbreitet</li> </ul> → <b>vorerst nicht nutzen (ggf. neu bewerten)</b> <ul style="list-style-type: none"> <li>• ggf. die normalen TIFF-Regeln beachten und die Georeferenz textlesbar dazu legen (z.B. als TFW-File, Esri-ArcGIS)</li> </ul>
*	SQL	<ul style="list-style-type: none"> <li>• standardisierte Datenbanksprache zur Definition von Datenstrukturen sowie zum Bearbeiten (auch Austauschen) und Abfragen von Daten</li> <li>• textlesbar</li> </ul>

\*\*\* = bevorzugtes Format, \* = akzeptiertes Format, ? = Kandidat (problematisch, ggf. aktuell neu bewerten)

## Anlage 2: Erstellung von Bildern für Bildarchive

(Sigmar Fitting)

Die klassische professionelle analoge Fotografie und die Archivierung der Silberhalogenidfotos stellte einen Prozess höchster Redundanz dar. Von jedem Bildmotiv wurden mindestens zwei Negative belichtet, sorgfältig entwickelt, archivfest fixiert und ausgewässert. Diese Negative wurden je nach Sicherheitsbedürfnis zusammen oder getrennt an verschiedenen Orten, als Rohmaterial archiviert und mit geeigneten Metadaten versehen. Das eigentliche Bild entstand daraufhin bei der Ausarbeitung im Labor, bei der durch die Auswahl des Ausschnittes und die Anpassung der Belichtung (Nachbelichtung, abwedeln, ...) die endgültige Bildaussage durch den Fotografen festgelegt wurde. In einem letzten Schritt erfolgte dann noch das Retuschieren, Aufziehen und Beschriften der Fotos. In aller Regel erfolgte darauf ein „Backup“ der fertigen Aufnahmen in Form eines Repros oder etwa durch die Mikroverfilmung (Bsp.: Marburger Index) der Archivkartons. Insgesamt also ein sehr aufwändiges Verfahren, welches die Notwendigkeit einer frühen Auslese der zu archivierenden Aufnahmen begründete.

Dieser klassische „Workflow“ lässt sich im Grunde sehr gut auf die Archivierung von digitalen Fotografien abbilden.

### 1. Rohdaten

Nach derzeitigem Stand und Kosten der Fototechnik sollten nur Aufnahmen akzeptieren die parallel als Rohdaten und ausgearbeitete Bilddatei abgeliefert werden. Je nach zur Verfügung stehendem Speicherplatz könnte es sehr sinnvoll sein diese Raw-Dateien als Belichtungsserien von mindestens drei Aufnahmen mit einer Schrittweite von 2 EV zu archivieren (damit eröffnen sich Möglichkeiten dem begrenzten Dynamikumfang von Digitalfotos etwa durch HDR/DRI Methoden zu begegnen). Diese RAW Dateien sollten dann sinnvoll benannt (Datenbank) und auf geeignete Weise mit Metadaten versehen werden und danach möglichst redundant an verschiedenen Orten und auf unterschiedlichen Speichermedien gesichert werden.

Die gleiche Vorgehensweise bietet sich bei Scans von Aufsichts- und Durchsichtsvorlagen an, wobei hier keine Belichtungsserien benötigt werden, da der abzubildende Dynamikbereich auch bei optimal belichteten Diapositiven leicht in einer 48 Bit RGB-Datei gespeichert werden kann – insbesondere im Bereich der Digitalisierung von Durchsichtsvorlagen bestimmt die eingesetzte Hardware ganz entscheidend die Qualität der Digitalisate (Dichteumfang, Schärfe).

#### *RAW-Dateiformate:*

Ein sehr hoffnungsvoller Ansatz bot 2004 die Einführung und Offenlegung des DNG-Formates als potentieller Standard für das „digitale Negativ“, dieser Standard konnte sich allerdings bis heute nicht so recht durchsetzen allerdings gibt es zum Beispiel mit Adobe Camera Raw ein Tool zum Konvertieren der proprietären RAW-Formate.

Nikon etwa hat bis heute noch nicht alle Spezifikationen des eigenen RAW-Formats offengelegt bei Canon verändern sich die Spezifikationen von einem Kameramodell zum Nächsten. Im HighEnd Bereich sieht die Sache nicht wirklich besser aus, hier bleibt die weitere Entwicklung abzuwarten und bis dahin (soweit alle Bearbeitungsmöglichkeiten erhalten bleiben sollen) die Kamera eigenen Raw-Dateien zu archivieren oder die Raw-Dateien in das DNG-Format zu konvergieren.

Im Scanbereich sieht es noch düsterer aus, etwa Imacon unterstützt in der neuesten Softwareversion das eigene „3f“-Format nicht mehr oder Vuescan-DNG's lassen sich nur mit Vuescan weiter bearbeiten, wenn man alle Bearbeitungsmöglichkeiten erhalten will. In diesem Bereich bietet sich eher das Speichern einer unkomprimierten 48-Bit Baseline-tiff-Datei an.

#### *TIFF*

An dieser Stelle einige Bemerkungen zum TIFF-Format. TIFF hat sich insbesondere im Bereich der Druckindustrie als Quasistandard für Rasterbilder durchgesetzt allerdings verfügt das TIFF-Format über außerordentlich viele Optionen und Varianten (Farbtiefe, Komprimierung, Bit-Anordnung, Ebenenkomprimierung, Farbkanäle, ...). Dies führt sehr schnell zu Inkompatibilitäten, ein Ausweg stellt



„Baseline-TIFF“ als Mindestspezifikation für alle „TIFF-fähigen“ Programme dar. TIFF-Dateien bieten darüber hinaus beste Möglichkeiten Metadaten (EXIF, IPTC und XMP) sowie ICC Profile in die Bilddatei zu integrieren.

#### *Empfehlungen zur Scan-Auflösung:*

Bei Farb- und Graustufen-Aufsichtsvorlagen hat man bei einer Auflösung von 600 dpi ausreichend Reserven, bei Bitmaps und Strichzeichnungen können je nach Vorlagengröße und gewünschtem Abbildungsmaßstab 1200 bis 2400 dpi sinnvoll sein. Bei Dia- und Negativfilmen erzielt man jenseits von 2400 bis 3000 dpi (bezogen auf die Negativgröße) schon eine „perfekte Abbildung des Filmkorns“.

## **2. Ausgearbeitete Bilddateien**

#### *Dateiformate:*

Für die Speicherung der fertigen Bilddateien gilt es eine Abwägung der sich teilweise widersprechenden Eigenschaften Speicherplatzbedarf <=> Qualität <=> Kompatibilität zu treffen. Eine sehr gute Grundlage bietet die oben skizzierte Archivierung der RAW-Dateien möglichst zusammen mit den entsprechenden „Kochrezepten“ für die „Entwicklung“ der fertigen Bilder. Diese Raw-Dateien bieten die maximale Qualität bei geringst möglicher Speichergröße der verlustfrei komprimierten Daten allerdings zum Preis einer sehr geringen Kompatibilität.

Unter der Voraussetzung, dass die endgültig ausgearbeiteten Bilddateien nicht weiter bearbeitet werden sollen stellt sich die Frage ob diese nicht als moderat komprimierte JPG Dateien, ggf. nach PDF/A-1b exportiert, archiviert werden sollten. Nach heutigem Stand bietet das JPG-Format (sämtliche Digitalkameras mit Ausnahme einiger High-End Digibacks speichern ausgearbeitete Bilddateien im JPG-Format) die größte Wahrscheinlichkeit dass diese Dateien auch in vielen Jahren ohne weitere Konvertierungen noch zu öffnen sein werden. Selbstverständlich sollten alle Zwischenstufen, soweit diese gespeichert werden sollen, nicht komprimiert werden.

Die Alternative mit der besten Kompatibilität stellt (Baseline-)TIFF (ggf. mit LZW-Komprimierung aber damit wird schon wieder die Kompatibilität eingeschränkt) zum Preis eines höheren Speicherplatzbedarfs dar.

#### *Auflösung und Farbtiefe:*

Nach heutigem Stand der Drucktechnik bietet eine Auflösung von 400 dpi reichlich Reserven, d.h. Für eine Abbildung in DIN A4 benötigt man eine Datei von ca. 3000\*4000 Pixeln für ein optimales Ergebnis und 2500\*3300 Pixel für ein gutes Druckergebnis. Bei der Farbtiefe reichen 8 Bit um den optimal möglichen Kontrastumfang eines Druckes auf weißem Papier abzudecken. Diese Bildgrößen entsprechen 12 bzw. 8,5 Megapixel bzw. einer Dateigröße von 36 bzw. 25,5 MByte pro unkomprimierter 8 Bit RGB-Datei.

#### *Farbkanäle und Farbprofil:*

Es ist sehr sinnvoll für die gesamte Bildbearbeitung und Archivierung einen RGB-Workflow beizubehalten und die Konvertierung in CMYK dem Druckertreiber, dem RIP-Programm oder der Druckerei zu überlassen. Sehr sinnvoll ist es natürlich ein ICC-Profil in die Datei einzubetten, als Standard bieten sich hier Adobe-RGB (von vielen Digitalkameras als Ausgabeprofil unterstützt) oder ECI-RGB (Standard im Grafischen Bereich) an (für die Ausarbeitung im Fotolabor sollte man die Dateien allerdings in das smallRGB Profil konvertieren)

## **3. Speicherung und Backup**

Ein wesentlicher Aspekt bei der Auswahl der zu archivierenden Dateien stellt die Dokumentation der verwendeten Formate und die Validierung der Dateien gemäß der standardisierten Spezifikationen. Ein dazu geeignetes Tool stellt etwa JHOVE dar. Zum Testen ist die von Harvard University entwickelte Version 1.x am besten geeignet, da sie auch über ein Testtool mit GUI verfügt ([http://sourceforge.net/projects/jhove/files/jhove/JHOVE%201.6/jhove-1\\_6.zip/download](http://sourceforge.net/projects/jhove/files/jhove/JHOVE%201.6/jhove-1_6.zip/download)).



Im Moment wird von verschiedenen Organisationen an Jhove 2.x gearbeitet (<https://bitbucket.org/jhove2/main/wiki/Home>):

*„Prinzipiell kann man nie genug über eine archivierte Datei wissen, jedoch kann es durchaus sinnvoll sein, extrahierte Metadaten einmal auf ihre Qualität zu überprüfen und gegebenenfalls für die Langzeitarchivierung nur indirekt relevante Daten herauszufiltern, um das Archivierungssystem nicht mit unnötigen Daten zu belasten. Beispiel für ein solches Tool ist „JHOVE“ (das JSTOR/Harvard Object Validation Environment der Harvard University Library, <http://hul.harvard.edu/jhove/>), mit dem sich auch Formaterkennung und Validierung durchführen lassen. Das Tool ist in Java geschrieben und lässt sich auch als Programmier-Bibliothek in eigene Anwendungen einbinden. Die generierten technischen Metadaten lassen sich sowohl in Standard-Textform, als auch in XML mit definiertem XML-Schema ausgeben.*

*... Validierungstools für Dateiformate stellen sicher, dass eine Datei, welche in einem fraglichen Format vorliegt, dessen Spezifikation auch vollkommen entspricht. Dies ist eine wichtige Voraussetzung für die Archivierung und die spätere Verwertung, Anwendung und Migration beziehungsweise Emulation dieser Datei. Das bereits erwähnte Tool „JHOVE“ kann in der aktuellen Version 1.1e die ihm bekannten Dateiformate validieren; verlässliche Validatoren existieren aber nicht für alle Dateiformate. Weit verbreitet und gut nutzbar sind beispielsweise XML Validatoren, die auch in XML Editoren wie „oXygen“ (SyncRO Soft Ltd., <http://www.oxygenxml.com>) oder „XMLSpy“ (Altova GmbH, <http://www.altova.com/XMLSpy>) integriert sein können.“*  
(Nestor Handbuch Ver. 2, Kap7/S. 16,17)

Eine sehr viel gröbere Erkennung bietet DROID auf Basis der PRONOM Format Registry der National Archives of Great Britain (<http://sourceforge.net/projects/droid/files/droid/> bzw. <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>)

Backups sollten grundsätzlich auf unterschiedlichen Medientypen erstellt werden. Denkbar wäre etwa ein Festplattensystem mit geeignetem Schreibschutz auf der einen Seite und die Ablage auf WORM-Bändern (LTO5, derzeit ca. 60 € / 1,5 TByte Speicherplatz und einer geschätzten Mindestlebensdauer von 50 Jahren) auf der anderen Seite. Dabei ist natürlich auch auf eine geeignete Speicherung und Archivierung der Informationen zu den verwandten Datentypen zu achten (Stand 10/2011).

Eine weitere sehr gute Möglichkeit wichtige Aufnahmen risikofrei für die nächsten 50 – 100 Jahre aufzubewahren stellt das Ausbelichten auf Fotopapier und aufbewahren in altmodischen Archivschränken dar. Genau so ernsthaft wird auch immer wieder die Ausbelichtung auf Dia- oder Mikrofilm diskutiert.