



Kurzfassung der IT-Empfehlungen für den nachhaltigen Umgang mit digitalen Daten in den Altertumswissenschaften

Herausgeber:
IANUS

Version 1.0.0.0
17. Juli 2017

Koordination



Förderung



Herausgeber IANUS

Autoren: siehe Autorenverzeichnis auf
<https://ianus-fdz.de/it-empfehlungen/autorenverzeichnis>

Titel: Kurzfassung der IT-Empfehlungen für den nachhaltigen Umgang mit digitalen Daten in den Altertumswissenschaften

Sprache: Deutsch

Zitierhinweis:

IANUS (Hrsg.), Kurzfassung der IT-Empfehlungen für den nachhaltigen Umgang mit digitalen Daten in den Altertumswissenschaften (2017) [Version 1.0.0.0]

Kontakt:

ianus-fdz@dainst.de
www.ianus-fdz.de

Lizenz:



Dieses Werk bzw. Inhalt ist lizenziert unter einer Creative Commons Namensnennung – Weitergabe unter gleichen Bedingungen 4.0 International Lizenz.

<http://creativecommons.org/licenses/by-sa/4.0/>

Versionshistorie

Version	Autor	Datum	Beschreibung
1.0.0.0	IANUS	15.05.2017	Erste veröffentlichte Kurzfassung.

Inhaltsverzeichnis

1	Einleitung	3
2	Projektphasen	4
2.1	Datenmanagement	5
2.1.1	Übersicht der Aufgaben in den Projektphasen	5
2.2	Dokumentation	7
2.2.1	Metadaten in der Anwendung	7
2.3	Dateiverwaltung	13
2.3.1	Dateiablage	13
2.4	Dateispeicherung und -sicherung	15
3	Dateiformate	16
3.1	PDF-Dokumente	17
3.2	Textdokumente	19
3.3	Bilder – Rastergrafiken	23
3.4	Bilder – Vektorgrafiken und CAD-Daten	26
3.5	Tabellen	31
3.6	Datenbanken	35
3.7	Video	42
3.8	Audio	48
3.9	3D und Virtual Reality	52
3.10	Webseiten	57
4	Weitere Informationen	63
4.1	Forschungsmethoden	63
4.2	Weitere Verweise	63

1 | Einleitung

E. Schneidenbach, F. Schäfer, M. Trognitz

Die Anwendung von Informationstechnologien und digitaler Methoden für die Durchführung archäologischer und altertumswissenschaftlicher Forschungsprojekte wird immer geläufiger und selbstverständlicher. Die von IANUS zusammengestellten IT-Empfehlungen sollen einen ersten Überblick über den nachhaltigen Umgang mit digitalen Forschungsdaten bieten und als Leitfaden dienen, um fachspezifische Mindeststandards im Bereich des Forschungsdatenmanagements zu etablieren. Eine konsequente Berücksichtigung dieser Empfehlungen trägt maßgeblich dazu bei, dass Forschungsdaten von Dritten, die nicht bei deren Erstellung und Verarbeitung beteiligt waren, heute und in Zukunft verstanden und nachgenutzt werden können.

Das vorliegende Dokument stellt die Kurzversion der online verfügbaren IT-Empfehlungen für den nachhaltigen Umgang mit digitalen Daten in den Altertumswissenschaften (doi:10.13149/000.111000-a) dar und konzentriert sich auf folgende Themen:

- Eine Übersicht der zentralen Aspekte im Datenmanagement
- Hinweise zur Dokumentation, Verwaltung und Sicherung von Daten
- Geeignete Langzeitformate und Metadaten für einzelne Dateitypen

Weitere Inhalte sind in der Online-Version zu finden:

- Ergänzende Hintergrundinformationen und praktische Tipps
- Hinweise zum Umgang mit in den Altertumswissenschaften angewandten digitalen Dokumentations- und Forschungsmethoden
- Archivierung von Forschungsdaten bei IANUS
- Glossar

Als allgemeine Empfehlungen lassen sich folgende Punkte zusammenfassen:

- Bei der Verarbeitung von Forschungsdaten ist deren Lebenszyklus zu beachten. Dieser umfasst die Erstellung, Verarbeitung, Analyse, Archivierung, den Zugang und die Nachnutzung. Diese Phasen sollten so früh wie möglich berücksichtigt werden und durch ein detailliertes, nachvollziehbares und gut strukturiertes Datenmanagement unterstützt werden. Hierfür ist eine solide Planung sowie eine konsequente Umsetzung erforderlich, die auch kontinuierliche Überprüfungen und eventuelle Anpassungen in der Praxis beinhaltet.
- Daten sollten während und nach der Erzeugung strukturiert und umfassend dokumentiert werden.
- Es sollten offen dokumentierte, nicht-proprietäre und weit verbreitete Dateiformate bevorzugt werden.
- Bei der Organisation von Dateien und Ordnern sollten möglichst einfache, klare und allen Akteuren bekannte Strategien vereinbart werden.
- Bei der Datensicherung sollten eindeutige Zuständigkeiten und Arbeitsabläufe definiert werden.

2 | Projektphasen

S. Jahn, F. Schäfer, M. Trognitz

Im Umgang mit digitalen Daten gelten für jede Phase eines Forschungsprojektes unterschiedliche Anforderungen und Bedingungen. Da Forschungsdaten einem Lebenszyklus unterliegen, haben Entscheidungen und Arbeitsschritte, die in einer bestimmten Phase getroffen werden, auch Auswirkungen auf die anderen Phase des Datenkreislaufs. Bereits bei Entwurf und Planung eines neuen Forschungsprojektes muss überlegt werden, welche Informationen bereits digital existieren, welche Arten von Dateien neu erzeugt werden müssen, welche Informationstechnologien zum Einsatz kommen sollen und wie das Management der Forschungsdaten gestaltet sein wird.

Insofern gilt es einige Aspekte so früh wie möglich, häufig noch vor der Erstellung der ersten Datei, zu adressieren:

- Informationen über existierende Standards und Richtlinien einholen
- Entscheidungen über die zu verwendenden Dateiformate und Softwareprogramme fällen
- Verantwortliche für das Datenmanagement bestimmen
- Kriterien für die laufende Dokumentation entwickeln
- Konventionen für die Ablage, Benennung und Versionierung von Dateien festlegen
- Strategien zur Speicherung und Sicherung definieren
- Geeignete Infrastrukturen für die Archivierung auswählen und kontaktieren

Diese und weitere Fragen zur Vorbereitung, Durchführung und Nachbereitung eines Projektes werden im Folgenden anhand eines sogenannten Datenmanagementplans thematisiert. Der Datenmanagementplan bildet einen wichtigen Baustein zur Festlegung der Arbeitsabläufe im Umgang mit Forschungsdaten und für eine Kostenkalkulation.

Je mehr und je früher in den Einzelentscheidungen die Nachnutzung der Daten durch Dritte auch über die Lebensdauer eines Projektes hinaus berücksichtigt wird und die Langzeitarchivierung als kontinuierliche Maßnahme und nicht als letzter Schritt eines bereits abgeschlossenen Projektes begriffen wird, umso leichter gelingt es, einmal erhobene Daten auch für die Zukunft zu erhalten. Die Archivierung der Daten nach Abschluss eines Projektes wird erleichtert, wenn schon während der Durchführung eines Projektes auf die konsequente Erfüllung der im Datenmanagementplan genannten Aufgaben geachtet wird.

Vor oder kurz nach Beginn des Projektes erstellte und kommunizierte Richtlinien, ermöglichen es allen Projektmitarbeitern diese auch einzuhalten. Auf diese Weise wird die Handhabung der Projektdaten erheblich erleichtert, da geregelte Ordnerstrukturen und vorgegebene Namenskonventionen das Auffinden der Daten vereinfachen, die parallele Dokumentation das Verstehen und die Nachnutzung ermöglicht und Sicherungskopien das Risiko eines Datenverlustes reduzieren.

2.1 Datenmanagement

M. Trognitz, D. Hagmann, J. Räther, S. Jahn

Bereits vor Beginn eines Forschungsvorhabens steht die Konzeption und Planung des Projektes. Dazu gehört auch eine Beschreibung über den Umgang mit den resultierenden Forschungsdaten, der sogenannte Datenmanagementplan. Ein vollständiger Datenmanagementplan berücksichtigt alle Phasen des auf Seite 6 beschriebenen Lebenszyklus von Forschungsdaten. Er dient zunächst als Mittel zur strukturierten Reflexion über datenrelevante Aspekte eines Projektes und beantwortet grundsätzliche Fragen zu Verantwortlichkeiten, Maßnahmen zur Pflege und Verarbeitung der Daten, Standards und bereits vorhandenen Daten. Außerdem bietet er die Grundlage für Arbeitsabläufe (Workflows) im Umgang mit Forschungsdaten und für eine Kostenkalkulation.

Eine vollständige Dokumentation des Datenmanagements in einem Datenmanagementplan spart Zeit und Kosten, wenn beispielsweise Zusammenhänge beim Wechsel von Mitarbeitern hergestellt werden sollen, und beugt einem Datenverlust vor. Da die Nachnutzung von Forschungsdaten zunehmend an Bedeutung gewinnt, setzen Geldgeber oft einen Datenmanagementplan als Teil eines Förderantrags voraus.

Die konsequente Einhaltung aller im Datenmanagementplan gemachten Vorgaben während der gesamten Projektlaufzeit stellt sicher, dass die Daten auch von Dritten interpretiert und nachgenutzt werden können. Wird das Überführen der Daten in ein Archiv schon von Beginn des Projektes an vorbereitet und einkalkuliert ist zum Projektabschluss nur noch ein geringer Aufwand für die Übergabe in ein Archiv erforderlich, weil die Umformatierung, Neustrukturierung oder nachträgliche Dokumentation von Daten wegfällt.

Ein aktives Datenmanagement beugt insbesondere während der Planung und Durchführung eines Projektes späteren Zeit- und Budgetverlusten vor und stellt sicher, dass Daten am Ende in nachhaltigen Formaten, gut dokumentiert und gut strukturiert vorliegen.

2.1.1 Übersicht der Aufgaben in den Projektphasen

Mit Hilfe des zu Beginn eines Projektes erstellten Datenmanagementplans können sämtliche Prozesse während des Projekts strategisch umgesetzt werden, wobei die Umsetzung des Plans als laufender Vorgang zu verstehen ist. Wenn während eines Projektes Änderungen an dem Datenmanagementplan notwendig werden, sollten diese begründet, dokumentiert sowie Arbeitsprozesse und Ergebnisse angepasst werden. Außerdem muss dokumentiert werden wie die Änderungen sich auf bereits bestehende Daten auswirken.

In den unterschiedlichen Phasen des Datenlebenszyklus und des Projektes sind die Aspekte des Datenmanagementplans unterschiedlich stark zu berücksichtigen. Die folgende Tabelle veranschaulicht, wann im Lebenszyklus von Forschungsdaten welchen Aufgaben aus dem Datenmanagement besondere Aufmerksamkeit zuteil werden muss.

In Abhängigkeit der Projektgröße kann der Umfang der Aufgaben des Datenmanagementplanes und der Plan selbst skaliert werden, wobei jedoch Minimalanforderungen einzuhalten sind, um zuallererst die Verfügbarkeit der Daten im laufenden Projekt zu gewährleisten. Um die Planungen zu erleichtern und

zu beschleunigen, sollten bereits vorhandene institutionelle Vorgaben und Infrastrukturen genutzt werden.

	Vorbereitung	Erstellung	Verarbeitung	Analyse	Archivierung	Zugang	Nachnutzung
Rahmendaten*	●	○	○	○	◐	○	○
Verantwortlichkeiten*	●	○	○	○	○	○	○
Rechtliche Aspekte*	●	○	○	○	●	●	●
Methoden	●	●	◐	◐	○	○	○
Vorgaben und Standards*	●	●	●	○	●	◐	○
Kosten & Ressourcen*	●	●	●	○	◐	○	○
Externe Partner	●	●	◐	○	○	◐	●
Hard- und Software*	●	●	◐	◐	○	○	○
Datentypen und Datenformate	◐	●	●	○	●	○	○
Nachnutzung vorhandener Daten*	◐	◐	●	●	○	◐	○
Datenerzeugung & -prozessierung*	◐	●	◐	○	◐	◐	○
Datenmenge	●	●	●	●	◐	○	○
Dateispeicherung und -sicherung*	◐	●	●	●	○	○	○
Dateiverwaltung	◐	●	●	●	◐	○	○
Dokumentation*	◐	●	●	●	●	◐	○
Qualitätssicherung	◐	●	●	●	●	○	●
Datenaustausch	◐	◐	●	●	○	●	○
Mittelfristige Datenaufbewahrung	◐	●	●	●	○	○	○
Langfristige Archivierung*	◐	◐	◐	◐	●	○	●
Zugänglichkeit und Nachnutzung	●	◐	◐	◐	●	●	●
Projektabschluss	◐	○	○	○	◐	◐	○

Tabelle 2.1: Tabellarische Übersicht über die verschiedenen zu berücksichtigenden Aspekte eines Datenmanagementplans während unterschiedlicher Projektphasen und unter Beachtung des Lebenszyklus von Forschungsdaten. Besonders wichtige, als Minimalanforderung zu betrachtende Aspekte des Datenmanagementplans sind mit einem Stern gekennzeichnet.

- Aufgabe ist während der Phase relevant.
- ◐ Aufgabe ist während der Phase teilweise relevant.
- Aufgabe ist während der Phase nicht relevant.

Das vollständige Kapitel finden Sie auf:
<https://ianus-fdz.de/it-empfehlungen/datenmanagement>

Den Lebensdatenzyklus von Forschungsdaten finden Sie auf:
<https://ianus-fdz.de/it-empfehlungen/lebenszyklus>

2.2 Dokumentation

S. Jahn

Neben technischen Aspekten ist eine vollständige Dokumentation das wichtigste Kriterium für die Archivierung und Nachnutzbarkeit von Daten. Damit Forschungsdaten von Dritten gefunden und sinnvoll verwertet werden können, müssen sie verständlich und strukturiert beschrieben werden. Ohne zusätzliche Informationen sind die meisten digitalen Dateien, die in einem Forschungsprojekt entstehen, für die Wissenschaft verloren und der Aufwand einer langfristigen Datenkuratierung ist vergeblich. Die Dokumentation und Beschreibung von Daten ist als ein kontinuierlicher Prozess zu begreifen, der in allen Phasen eines Datenlebenszyklus adressiert werden sollte.

Metadaten können sich auf folgende Aspekte beziehen:

- Deskriptive Angaben
- Strukturelle Angaben
- Administrativ-rechtliche Angaben
- Administrativ-erhaltungsbezogene Angaben
- Technische Angaben

Um Metadaten eindeutig zu vergeben und eine maschinelle Verarbeitung zu ermöglichen, sollten kontrollierte Vokabulare in Form von einfachen Wertelisten oder strukturierten Thesauri verwendet werden. Auch sollten standardisierte und etablierte Metadatenschemata, wie beispielsweise Dublin Core benutzt werden.

Die Speicherung der Metadaten kann teilweise innerhalb der Datei selbst erfolgen, wobei die Speicherung der Informationen in einer eigenen Datei, wie etwa einer Fotoliste, empfohlen wird.

2.2.1 Metadaten in der Anwendung

Projektbezogene Metadaten

Die wichtigsten Metadaten, die für die Beschreibung eines Projektes oder einer Dokumentensammlung erforderlich sind, werden in der folgenden Tabelle abgebildet und knapp definiert. Sie geben einen Überblick über einen größeren, zusammenhängenden Datenbestand und beschreiben den fachlichen Kontext in dem dieser entstanden ist. Vergleichbar einem Bibliothekskatalog liegt die Hauptfunktion dieser Metadaten darin, dass externe Personen ein Projekt oder eine Dokumentensammlung über Web-Portale und Suchmaschinen finden und einordnen können. Darüber hinaus enthalten sie rechtliche Informationen, die für den weiteren Umgang mit den Daten wichtig sind.

Die hier vergestellten Eigenschaften basieren auf dem Dublin Core Metadata Schema und den Angaben, wie sie vom ADS in UK und tDAR in USA erhoben werden, um einen zukünftigen Austausch zu vereinfachen. Ein darauf aufbauendes ausführliches Metadatenschema, das auch die Grundlage für die Archivierung von Projektdaten bei IANUS bildet, sowie ausgefüllte Beispielformulare sind in dem Kapitel Archivierung von Forschungsdaten in IANUS ab Seite 12 zu finden.

Bezeichnung	Kurzdefinition
Identifizierung – Projekttitel	Verbindliche Kurzbezeichnung des Projektes.
Identifizierung – Alternativtitel	Ggf. alternative Titel für ein Projekt.
Identifizierung – Projektnummer(n)	Nummern oder Kennungen, die z.B. innerhalb der durchführenden Organisation oder von Mittelgebern verwendet wird, um das Projekt eindeutig identifizieren zu können.
Kurzbeschreibung	Knappe Angaben zur Fragestellung, zum Verlauf und Ergebnis des Projektes sowie Skizzierung der Datensammlung (insgesamt ca. 100-300 Worte)
Schlagworte – Fachdisziplinen	Stichworte, die die beteiligten Disziplinen und Fächer benennen. Sofern die Stichworte auf publizierten Standards oder internen Thesauri beruhen, müssen diese mitangegeben werden.
Schlagworte – Inhalt	Stichworte, die den Inhalt der Datensammlung benennen., z. B. zu Materialgruppen, Fundstellen-Klassifizierung, Quellenarten, Kulturgruppen etc. Sofern die Stichworte auf publizierten Standards oder internen Thesauri beruhen, müssen diese mitangegeben werden.
Schlagworte – Methoden	Stichworte, die die eingesetzten Forschungsmethoden beschreiben. Sofern die Stichworte auf publizierten Standards oder internen Thesauri beruhen, müssen diese mitangegeben werden.
Ausdehnung – Geografisch-1	Detaillierte Angaben zur räumlichen Ausdehnung oder zum Fundort des untersuchten Gegenstandes mittels geografischer Koordinaten. Die maximale Ausdehnung kann als Bounding Box angegeben werden.
Ausdehnung – Geografisch-2	Sprachliche Beschreibung des untersuchten Gegenstandes mittels Ortsangaben mit Land, Stadt, Kreis, Straße, Gemarkung etc. Sofern Namen sich im Lauf der Zeit geändert haben, dies gesondert vermerken. Sofern eine Referenz zu einer Geo-Ressource oder einem Gazetteer existiert, sollte diese ebenfalls angegeben werden.
Ausdehnung – zeitlich	Chronologische Angaben zum untersuchten Gegenstand, entweder als Periodenbezeichnung und/oder mit groben/genauen Jahresangaben. Sofern die Stichworte auf publizierten Standards oder internen Thesauri beruhen, müssen diese mitangegeben werden.

Fortsetzung auf der nächsten Seite

Bezeichnung	Kurzdefinition
Primärforscher – Person	Personen, die entweder für das Projekt als Ganzes, für das Datenmanagement oder für die Erzeugung bestimmter Datenarten zentral bzw. verantwortlich sind. Hier ist eine Kontaktadressen erforderlich und die aktuelle/letzte institutionelle Zugehörigkeit, damit die Personen bei Rückfragen erreicht werden kann.
Eigentümer – Organisation	Organisation, der die unter „Primärforscher“ genannten Personen angehören, oder die nach Ausscheiden derselben für die Daten verantwortlich ist, im weitesten Sinne also Eigentümer der Daten ist. Hier ist eine Kontaktadresse erforderlich, damit die Organisation bei Rückfragen erreicht werden kann.
Finanzierung	Nennung der Organisation(en) / (Dritt-)Mittelgeber, durch die das Projekt finanziert wurde. Es sollte jeweils der Zeitraum der Finanzierung angegeben werden.
Veröffentlichung – Projektdaten	Wenn die hier beschriebene Datensammlung des Projektes bereits an anderer Stelle veröffentlicht / online gestellt wurde, bitte entsprechende Angaben machen, z. B. durch Nennung der Organisationen, Datenarchive, Online-Ressourcen etc.
Veröffentlichung – Ergebnisse	Analoge oder digitale Publikationen zu Ergebnissen des Projektes oder zur Datensammlung des Projektes, ausführliche bibliographische Angaben (ohne fachspezifische Abkürzungen) unter Nennung des Verlages erforderlich.
Dauer – Projekt	Anfangs- und Enddatum des Projektes.
Dauer – Datenbestand	Anfangs- und Enddatum der Erzeugung oder Verarbeitung digitaler Daten im Rahmen des Projektes.
Rechtliches – Urheberrechte	Name des Inhabers der Urheber-, Nutzungs- und Verwertungsrechte; i. d. R. die Organisation, an der der Primärforscher beschäftigt war.
Rechtliches – Lizenzgeber	Angabe der Person, die i. d. R. als Vertretung für eine Organisation für die Lizenzierung von Daten zur Nachnutzung verantwortlich und berechtigt ist, einen Datenübergabevertrag abzuschließen.
Rechtliches – Datenschutz	Angaben, ob in der Datensammlung datenschutzrelevante Informationen enthalten sind. Wenn ja, in welchem Umfang.
Quellen – Ältere	Ältere Quellen oder existierende Ressourcen, auf denen die Daten aufbauen.

Bezeichnung	Kurzdefinition
Quellen – Zugehörige	Sofern während des Projektes Informationen, Datensammlungen, (un-)publizierte Dokumente, Online-Ressourcen etc. verwendet oder erzeugt wurden, die nicht Teil der hier beschriebenen Datensammlung sind, aber für deren Verständnis wichtig sind, bitte entsprechende Angaben zu Art und Umfang dieser Quellen machen.
Sprache	Die in den Dokumenten und Dateien verwendete(n) Sprache(n). Sprachkennungen nach ISO 639 angeben.
Art der Daten	Kurzcharakterisierung der Daten, z. B. ob es sich um Rohdaten, verarbeitete Daten, Interpretationen, Ergebnisse, Abschlussberichte etc. handelt.
Vollständigkeit	Aussagen zur Vollständigkeit der Projektdaten, z. B. ob bestimmte Datenarten noch fehlen und warum.
Dateiformate	Auflistung der Dateiformate, die in der Datensammlung vorkommen, ggf. unter Nennung der verwendeten Programme und Zeichenkodierungen.
Zugriffsrechte	Festlegung der gewünschten Zugriffsrechte für die Daten, sofern diese für den gesamten Projekt-Datenbestand gelten sollen; differenzierte Regelungen müssen auf Dateiebene vorgenommen werden.
Signatur Metadaten	Angabe darüber, wer die o. g. Metadaten wann ausgefüllt hat.

Dateibezogene Metadaten

Bei dieser Art von Metadaten handelt es sich um technische und inhaltliche Informationen, die Nutzern verständlich machen, wie einzelne Dateien innerhalb eines Projektes oder einer Datensammlung beschaffen sind und welche Möglichkeiten der Nachnutzbarkeit sie beinhalten.

Dateibezogene Metadaten sind abhängig von dem Dateiformat, dem Inhalt und der Methode, mit der sie erzeugt wurden. Beispielsweise sind für Rastergrafiken, die durch digitale Fotografie entstanden sind, andere Angaben erforderlich (Fotograf, Aufnahmedatum, Aufnahmeort, abgelichtetes Objekt etc.) als für Rasterdateien, die durch geophysikalische Messungen erzeugt wurden (Koordinaten, Messgerät, Genauigkeit, Datum etc.). Zusätzliche spezifische Angaben, die für bestimmte Dateiformate empfohlen werden, sind in den verschiedenen Kapiteln zu den jeweiligen Formaten ab Seite 16 beschrieben.

Es gibt jedoch dateibezogene Metadaten, die unabhängig von Format, Inhalt und Methode für alle Einzeldateien gleichermaßen relevant und notwendig sind. Zu diesen Metadaten gehören neben den Angaben zu Dateiname, Dateiformat, Dateiversion, Titel, Beschreibung und Ersteller auch detaillierte Angaben zur verwendeten Soft- und Hardware, Versionierung, rechtliche Aspekte und weitere abhängige Dateien.

Auch wenn es in der Theorie wünschenswert ist, diese Angaben sowie die zugehörigen spezifischen Angaben zu Methoden und Dateiformaten für jede Datei einzeln zu erfassen, so zeigt die Praxis, dass es häufig ausreichend ist, einen Metadatensatz für Gruppen von Dateien anzulegen, wenn diese das gleiche Format oder die gleichen inhaltlichen Eigenschaften aufweisen.

Bezeichnung	Kurzdefinition
Identifikator, Dateiname	Eindeutiger Name der Datei.
Dateiformat	Format, in der die Datei abgespeichert ist.
Urheber	Name des Verfassers oder Erstellers der Datei; gegebenenfalls Name der Einrichtung.
Titel	Titel der Datei, nicht der Dateiname.
Beschreibung	Beschreibung des Inhalts der Datei.
Schlagworte	Schlagworte, wie etwa Periode, Fundstelle oder charakteristische Merkmale. Wenn vorhanden, angemessene Thesauri verwenden.
Software – Dateierstellung	Software, mit der die Datei erstellt wurde.
Hardware – Dateierstellung	Hardware, mit der die Datei erstellt wurde, v. a. bei technischen Geräten wie Kameras, GPS-Geräten, Vermessungsinstrumenten, Laserscanner etc.
Betriebssystem – Dateierstellung	Betriebssystem, das verwendet wurde als die Datei erstellt wurde.
Erstellungsdatum	Datum, an dem die Datei erstellt wurde. Datum und Zeit in UTC nach ISO 8601.
Letzte Aktualisierung	Datum, an dem die Datei zuletzt bearbeitet wurde. Datum und Zeit in UTC nach ISO 8601.
Dateiversion	Angabe der Versionsnummer der Datei.
Weitere Dateien	Referenzen auf Dateien, die für das Verständnis einer anderen Datei zentral sind, insbesondere für zusammenhängende, komplexe Dateien oder wenn auf eine Ursprungsdatei verwiesen werden soll.
Sprache	Sofern schriftliche Inhalte vorhanden sind, die Sprache angeben. Sprachkennungen nach ISO 639 angeben.
Copyright-Angaben	Angaben zur Person oder Einrichtung, die das Copyright oder die Lizenzrechte an der Datei oder deren Inhalt besitzt.

Methodenbezogene Metadaten

Jede Fachdisziplin innerhalb der Altertumswissenschaften verfügt über eigene Forschungsmethoden. Auch diese haben Einfluss auf Umfang, Art, Struktur und Inhalt digitaler Objekte, da sie bei verschiedenen Arbeitsweisen und technischen Geräten unterschiedlich ausfallen. Daher sollten methodenbezogene Metadaten ebenfalls dokumentiert werden, insbesondere wenn mehrere Zwischenstände einer Prozesskette archiviert und Dritten zur Verfügung gestellt werden sollen.

Je nach Genauigkeit der Methodenbeschreibung kann sie sich sowohl auf eine Einzeldatei als auch auf mehrere Dateien gleichen Typs beziehen.

Zusätzliche spezifische Angaben, die für bestimmte Forschungsmethoden oder Prozesse empfohlen werden, sind in den einzelnen Kapiteln des Abschnittes „Forschungsmethoden“ ab Seite 63 beschrieben.

Bezeichnung	Kurzdefinition
Prozessnummer	Eindeutige Nummer eines Prozesses oder einer Methode.
Prozessbeschreibung	Beschreibung des Prozesses/der Methode, vor allem der Ausgangssituation und der Zielvorstellung.
Ausgangsformat(e)	Format der Dateien, die am Anfang eines gesamten Prozesses stehen und den Ausgangspunkt bilden.
Zwischenformat(e)	Format der Dateien, die im Verlauf eines Prozesses erzeugt werden und den Ausgangspunkt für weitere Prozesse bilden.
Zielformat(e)	Format der Dateien, die am Ende eines gesamten Prozesses erzeugt werden und den Zielpunkt definieren.
Durchführender	Person(en), die den Prozess durchgeführt hat (haben).
Prozessbeginn	Datum, an dem der Prozess begonnen wurde. Datum und Zeit in UTC nach ISO 8601.
Prozessende	Datum, an dem der Prozess beendet wurde. Datum und Zeit in UTC nach ISO 8601.
Software	Software, mit der der Prozess durchgeführt wurde.
Hardware	Hardware, auf der der Prozess durchgeführt wurde.

Das vollständige Kapitel finden Sie auf:

<https://ianus-fdz.de/it-empfehlungen/dokumentation>

Hinweise zur Archivierung bei IANUS finden Sie auf:

<https://ianus-fdz.de/it-empfehlungen/archivierung>

2.3 Dateiverwaltung

M. Trognitz

Der tägliche Umgang mit digitalen Daten wird durch eine effiziente Dateiverwaltung erheblich erleichtert. Aussagekräftige Dateinamen, die auch für Dritte verständlich sind, sorgen dafür, dass die Dateien gefunden und deren Inhalt verstanden wird. Einheitliche Dateinamensstrukturen erhöhen die Lesbarkeit. Die konsequente Einhaltung von Versionierungsangaben sorgt dafür, dass immer mit der richtigen Dateiversion gearbeitet wird und eine selbsterklärende Ordnerstruktur hilft dabei auch in großen Projekten bestimmte Dateien wiederzufinden.

Die Konzipierung der Dateiverwaltung muss schon zu Beginn des Projektes erfolgen, damit die Regeln von Anfang an angewendet werden können. Die Niederschrift der durch Beispiele ergänzten Benennungsregeln und deren Weitergabe dient im laufenden Betrieb als wichtiges Nachschlagewerk für die konsistente und konsequente Einhaltung derselben. Dies ermöglicht dann eine effizientere Arbeitsweise.

Im laufenden Projektbetrieb sollte die Einhaltung der Benennungsregeln kontrolliert und gegebenenfalls angepasst werden.

2.3.1 Dateiablage

Mit Dateiablage ist hier vor allem die Ordnerstruktur gemeint. Für die Benennung der Ordner gelten die gleichen Regeln, wie für die Dateibenennung, die im nächsten Abschnitt thematisiert werden. Lediglich die Dateinamenserweiterung wird bei Ordnernamen nicht verwendet. Die Dateiablage sollte selbsterklärend sein und unpräzise Namen wie etwa „*In Arbeit*“ sollten unbedingt vermieden werden.

Wichtig ist, dass die Dateiablage logisch und hierarchisch aufgebaut ist. Dies bedeutet unter anderem, dass die Inhalte und Unterordner auch tatsächlich thematisch in den übergeordneten Ordner hineinpassen. Beispielsweise erwartet man in einem Ordner mit dem Namen „*Fotos*“ diverse Fotos, die bei einer entsprechenden Menge vielleicht noch auf verschiedene Unterordner, wie zum Beispiel „*Plana*“ und „*Profile*“, verteilt sind.

Ist aber in dem Ordner „*Fotos*“ ein Unterordner mit dem Namen „*Zeichnungen*“ enthalten, in dem verschiedene digitalisierte Zeichnungen abgelegt sind, führt das zu Schwierigkeiten. Der Unterordner wird wahrscheinlich nur schwer wiedergefunden werden, da der Name des übergeordneten Ordners einen anderen Inhalt verspricht.

Tritt der Fall ein, dass eine Datei oder ein Ordner thematisch in mehrere verschiedene übergeordnete Ordner passen würde, können verschiedene Lösungswege zur Anwendung kommen. Beispielsweise könnte eine Kopie abgelegt werden, was allerdings problematisch ist, wenn eine der Kopien verändert wird und diese nicht mit der anderen abgeglichen wird.

Alternativ könnte eine Dateiverknüpfung angelegt werden. Dateiverknüpfungen verschiedener Betriebssysteme sind meist nicht zu anderen Systemen kompatibel und funktionieren im schlimmsten Fall sogar nur auf dem einen Rechner, auf dem sie erstellt wurden. Dies liegt zum einen daran, dass jedes Betriebssystem unterschiedliche Methoden zur Erstellung verwendet und zum anderen können schon die Laufwerke unterschiedlich benannt sein. Außerdem

kann die Dateiverknüpfung ins Leere führen, wenn die originale Datei verschoben oder gelöscht wurde. Daher ist von Dateiverknüpfungen abzuraten.

Eine bessere Methode ist das Anlegen einer Textdatei mit einem Hinweis auf den Ablageort der Datei oder des Ordners. In diesem Fall muss aber darauf geachtet werden, dass die gesuchte Datei sich auch tatsächlich an dem angegebenen Ort befindet. Bei Verschieben oder Löschen müsste die Textdatei also auch immer berücksichtigt und angepasst werden.

Am besten ist es, nur eine Instanz einer Datei oder eines Ordners zu haben. Bei Zuordnungsschwierigkeiten kann es helfen, die Datei oder den Ordner in der hierarchischen Struktur höher anzusiedeln.

Eine Textdatei kann auch verwendet werden, um die vorhandene Ordnerstruktur zu erklären oder auf Besonderheiten hinzuweisen. Beispielsweise kann in einem Ordner mit Fotos, deren Metadaten in einer Datenbank abgespeichert sind, mit Hilfe einer Textdatei verdeutlicht werden, wo die Metadaten zu finden sind. Damit die Textdatei auch als eine Hilfedatei erkannt wird, kann man ihr den Namen „*README*“ oder „*LIESMICH*“ geben. Die Schreibung mit Großbuchstaben erhöht die Auffälligkeit der Datei. Durch eine vorangestellte Null kann sie dann auch in der alphabetischen Sortierung ganz an den Anfang gestellt werden.

Das vollständige Kapitel finden Sie auf:

<https://ianus-fdz.de/it-empfehlungen/dateiverwaltung>

2.4 Dateispeicherung und -sicherung

M. Trognitz, R. Komp, R. Förtsch

Um Datenverlust vorzubeugen, ist es unerlässlich eine geeignete Sicherungsstrategie zu verwenden. Dabei muss zwischen einer kurzfristigen Speicherung und einer mittelfristigen Sicherung unterschieden werden. Ersteres meint, wie Daten während der Erstellung und Bearbeitung gespeichert werden. Letzteres bezieht sich auf einen längeren Speicherzeitraum, der durchaus auch mehrere Monate betragen kann, stellt also ein klassisches Backup der Daten dar.

Die Sicherungsstrategie legt fest, wie die Datensicherung erfolgen soll und berücksichtigt folgende Fragen:

- Wer ist für die Datensicherung verantwortlich?
- Wer hat Zugriff auf die gesicherten Daten?
- Wann und wie oft soll die Datensicherung durchgeführt werden?
- Auf welche Weise soll gesichert werden?
- Welche Daten sollen gesichert werden?
- Welche Speichermedien sollen verwendet werden?
- Wie viele Sicherungskopien sollen angelegt werden?
- Wo sollen die Sicherungen aufbewahrt und wie sollen sie geschützt werden?
- Wie soll der Transport der Sicherungskopien erfolgen?
- Wie lange soll eine Sicherung aufbewahrt werden?

Die Sicherungsstrategie kann mit den zu verwendenden Richtlinien zur Dateiablage eng verzahnt sein, um etwa einen nahezu automatisierten Sicherungsvorgang zu ermöglichen.

Die richtige Speicherstrategie beugt zwar möglichen Datenverlusten durch Hardware-, Software- oder menschliche Fehler vor, stellt aber noch *keine* Archivierung der Daten dar. Eine Archivierung ist auf Langfristigkeit ausgelegt und impliziert immer eine bewusste Auswahl und umfassende Dokumentation der Daten, da eine Nachnutzung derselben das Ziel ist. Während der Projektlaufzeit kann bereits ein eigener Archivordner angelegt werden, worin beispielsweise finale Dateien oder unprozessierte Rohdaten abgelegt werden können, um die spätere Auswahl der zu archivierenden Daten vorzubereiten und zu vereinfachen.

Weiterführende Hinweise zur Datensicherung sind auf den Seiten des Bundesamtes für Sicherheit in der Informationstechnik zu finden, die sich sowohl an Einsteiger¹ als auch an Experten² richten.

Das vollständige Kapitel finden Sie auf:

<https://ianus-fdz.de/it-empfehlungen/dateispeicherung>

¹https://www.bsi-fuer-buerger.de/BSIFB/DE/MeinPC/Datensicherung/Sicherungsmethoden/sicherungsmethoden_node.html

²https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/_content/baust/b01/b01004.html

3 | Dateiformate

F. Schäfer

Im Folgenden werden konkrete technische Hinweise und Hintergrundinformationen zu den Dateiformaten gegeben, die nach heutigem Wissen für die Langzeitarchivierung geeignet sind. Des Weiteren werden die minimalen Anforderungen an die technische wie inhaltliche Dokumentation von einzelnen Dateiformaten genannt, damit diese nicht nur heute, sondern auch in Zukunft von Dritten verstanden und nachgenutzt werden können.

Es geht hier nur um die Dateiformate an sich, ihre Entstehungsweise wird dabei nicht berücksichtigt. Abhängig von der jeweiligen Generierung und Verwendung sind gegebenenfalls zusätzliche methodische und fachspezifische Metadaten erforderlich. Diese werden in dem Kapitel Forschungsmethoden auf www.ianusfdz.de/it-empfehlungen/forschungsmethoden beschrieben.

Bereits die Auswahl der verwendeten Software beeinflusst das Dateiformat und dessen Nachhaltigkeit. Daher sollte bei der Entscheidung für oder gegen eine bestimmte Software darauf geachtet werden, dass Dateiformate unterstützt werden, die möglichst folgende Eigenschaften aufweisen:

- Weit verbreitet und standardisiert
- Nicht proprietär, also nicht von einer Anwendung oder einem Hersteller abhängig und mit unterschiedlichen Programmen verwendbar
- Offen dokumentiert mit frei verfügbaren technischen Spezifikationen
- Verlustfreie oder keine Kompression
- Einfach dekodierbar oder sogar unmittelbar lesbar, also nicht durch Kodierung versteckt

Werden während eines Arbeitsprozesses andere als die hier empfohlenen Dateiformate erzeugt oder verwendet, sollte darauf geachtet werden, dass diese leicht und mit möglichst geringem Verlust an Informationen und Funktion in eines der in den folgenden Kapiteln aufgeführten Langzeitformate überführt werden können.

In den Übersichtstabellen mit den Formaten für bestimmte Dateitypen sind die präferierten Formate mit einem grünen Haken, die noch akzeptablen mit einer gelben Tilde und für die Langzeitarchivierung ungeeignete Formate mit einem roten Kreuz gekennzeichnet.

3.1 PDF-Dokumente

S. Jahn; Mit Unterstützung von: D. von Seggern

Das PDF (Portable Document Format) wurde 1993 von Adobe Systems entwickelt, um den Datenaustausch zu erleichtern. Es ist ein plattformunabhängiges, offenes Dateiformat, das 2008 mit der Version 1.7 als ISO-Standard zertifiziert wurde und seitdem von der ISO weiter gepflegt wird.

Der große Vorteil von PDFs liegt darin, dass Dateiinhalte unabhängig vom Betriebssystem, dem ursprünglichen Anwendungsprogramm und der Hardwareplattform unverändert dargestellt werden. Das Aussehen von Dokumenten wird, wie bei einem analogen Ausdruck, eingefroren und kann somit angezeigt werden, wie es ursprünglich vom Autor intendiert war. Gleichzeitig sind die Möglichkeiten zur nachträglichen Bearbeitung begrenzt, wodurch eine große Authentizität gewährleistet wird.

Für das Öffnen einer PDF-Datei gibt es verschiedene freie Anwendungen, worin ein Grund für die große Verbreitung und Akzeptanz von PDFs liegt. Viele Programme können Dateien direkt im PDF-Format speichern oder exportieren. Darüber hinaus lassen sich mit Hilfe von zusätzlich installierten Druckertreibern PDF-Dateien aus allen Programmen heraus erzeugen.

Langzeitformate Nicht jede Datei mit der Dateierweiterung .pdf ist gleichermaßen für die Langzeitarchivierung geeignet. Für diesen Zweck wurde das PDF/A-Format entwickelt und als ISO-Standard zertifiziert. Es beschreibt in welcher Form bestimmte Elemente in einer PDF-Datei enthalten sein müssen und welche nicht erlaubt sind. Die erste Version der PDF/A-Norm wurde nachträglich um zwei weitere, aufeinander aufbauende Normteile ergänzt. Wird das PDF/A-Format für die Langzeitarchivierung einer Datei verwendet, die nicht nur textuelle Informationen enthält, sollten nach Möglichkeit auch die ursprünglichen Ausgangsdateien in archivierungstauglichen Formaten entweder als separate Dateien oder integriert in das Container-Format PDF/A-3 mitarchiviert werden.

Format		Begründung
✓	PDF/A-1	PDF/A ist gezielt als stabiles, offenes und standardisiertes Format für die Langzeitarchivierung unterschiedlicher Ausgangsdateien entwickelt worden.
	PDF/A-2	
~	PDF/A-3	PDF/A-3 ist nur dann für die Langzeitarchivierung geeignet, wenn alle eingebetteten Dateien in einem anerkannten Archivformat vorliegen.
✗	andere PDF-Varianten	Viele gängige PDF-Varianten sind nicht für die Langzeitarchivierung geeignet. Stattdessen sollten entweder die Ausgangsdateien in einem passenden Format archiviert oder eine Migration in ein PDF/A-Format vorgenommen werden.

Dokumentation Welche Metadaten für ein PDF-Dokument relevant sind, hängt von dessen Inhalt ab. Bei PDF-Dateien, die nur Text enthalten, sind meist weniger Informationen erforderlich, als bei PDFs, die Bilder, GIS-Karten oder 3D-Modelle enthalten. Die Metadaten können unmittelbar in einem PDF-Dokument im XMP-Format gespeichert werden.

Die hier angegebenen Metadaten sind als minimale Angabe zu betrachten und ergänzen die angegebenen Metadaten für Projekte und Einzeldateien in dem Abschnitt Metadaten in der Anwendung ab Seite 7.

Metadatum	Beschreibung
Titel	Titel des Dokuments, nicht der Dateiname
Autor	Name des Verfassers oder Erstellers der Datei; gegebenfalls Name der Einrichtung
Stichwörter	Schlagworte, wie z.B. Periode, Fundstelle oder charakteristische Merkmale. Wenn vorhanden, angemessene Thesauri verwenden
Beziehungen	Dateien oder Ressourcen, die mit dem PDF zusammenhängen, insbesondere der Name der Originaldatei, aus der heraus ein PDF erstellt wurde
Anwendung	Programm, mit dem der Inhalt ursprünglich erzeugt wurde
Datum	Datum der Erstellung oder letzten Änderung der Datei
Copyright-Angaben	Angaben zur Person oder Einrichtung, die das Copyright oder die Lizenzrechte an der Datei oder deren Inhalt besitzt
Zusätzliche Metadaten	
Kurzbeschreibung	Kurzbeschreibung über den Inhalt des Dokuments
Sprache	Sofern schriftliche Inhalte vorhanden sind, die Sprache angeben. Sprachkennungen nach ISO 639 angeben.
Verfasser der Metadaten	Name der Person, welche die Metadaten ausgefüllt hat

Weitere Metadaten sind abhängig vom Inhalt und der Methoden und können in den jeweiligen Abschnitten im Kapitel Forschungsmethoden ab Seite 63 nachgelesen werden.

Das vollständige Kapitel finden Sie auf:

<https://ianus-fdz.de/it-empfehlungen/pdf-dokumente>

3.2 Textdokumente

M. Trognitz

Textdokumente stellen in der altertumswissenschaftlichen Forschung einen häufig vertretenen Dateityp dar. In Artikeln, Berichten, Anträgen, Tagebüchern, Notizen oder Dokumentationen sind wichtige Informationen enthalten, deren Fortbestehen und Lesbarkeit gewährleistet sein müssen. Auch Beschreibungen von anderen Dateien und Datensätzen oder des gesamten Projektes können als Textdokumente vorliegen.

Die Mehrzahl der Dokumente besteht aus strukturiertem Text, nämlich Sätzen, Absätzen, Seiten, Fußnoten und Kapiteln, und kann Formatierungsangaben, wie verschiedene Schriftgrößen, Fett- oder Kursivschreibung enthalten. Zusätzlich können Medien, wie Bilder, Tabellen oder Videos in die Dokumente integriert sein.

Da dasselbe Dokument auf verschiedenen Systemen unterschiedlich dargestellt werden kann, kann die Speicherung von Textdokumenten problematisch sein. Insbesondere wenn bestimmte Formatierungen von Textelementen mit einer Bedeutung verbunden sind und die Authentizität des Erscheinungsbildes, also das Aussehen des Dokumentes wichtig ist, ist bei der Speicherung besondere Aufmerksamkeit erforderlich.

Langzeitformate Finalisierte Dokumente mit Formatierungsangaben können im Format PDF/A gespeichert werden. Dieses Format erlaubt eine konsistente Darstellung des Dokumentes auf verschiedenen Systemen, verhindert aber auch eine nachträgliche Bearbeitung. Nähere Informationen sind im Abschnitt über PDF-Dokumente ab Seite 17 zu finden.


Textdokumente mit Formatierungsangaben, bei denen auch weiterhin eine Bearbeitung möglich sein soll, sollten in einem offenen auf XML basierenden Format gespeichert werden, wie beispielsweise DOCX oder ODT. Ersteres ist das Standardformat, das in Microsoft Word seit 2007 verwendet wird und auch von Microsoft entwickelt wurde. Letzteres ist das Format für Textdokumente, welches in OpenOffice oder LibreOffice verwendet wird. ODT ist ein Teil vom OpenDocument Format (ODF) und wurde von einem technischen Komitee unter der Leitung der *Organization for the Advancement of Structured Information Standards* (OASIS) entwickelt. Die Darstellung von DOCX- oder ODT-Dokumenten kann jedoch von System zu System unterschiedlich ausfallen, wenn beispielsweise bestimmte Schriftarten fehlen. Gegebenfalls kann das Dokument parallel im Format PDF/A gespeichert werden.

Bearbeitbare Textdokumente ohne Formatierungsangaben werden am besten als TXT-Datei gespeichert. Neben diesem einfachen, reinen Textformat (*plain text*) gibt es weitere textbasierte Formate, die auf eine bestimmte Weise strukturiert sind oder eine Auszeichnungssprache verwenden. Es handelt sich dabei um sogenannte Textdateien, die im Gegensatz zu binären Formaten darstellbare Zeichen enthalten und in Abhängigkeit ihrer Strukturierung unterschiedliche Dateiformate beschreiben. Beispielsweise werden mit Hilfe von CSV-Dateien Tabellen oder mit PLY-Dateien 3D-Inhalte gespeichert. Diese Formate werden in den Abschnitten „Tabellen“ und „3D und Virtual Reality“ ab Seite 31 bzw. 52 behandelt. Für textuelle Inhalte gibt es spezialisierte Formate, wie beispielsweise SGML-, XML- oder HTML-Dateien. Die Archivierung dieser Dateien, die

weit verbreiteten Konventionen folgen, ist unproblematisch, bedarf jedoch zusätzlicher Dateien, die die verwendete Struktur beschreiben, wie beispielsweise die sogenannte Dokumenttypdefinition (DTD, Document Type Definition) oder ein XML Schema (XSD, XML Schema Definition). Auch andere Textdateien mit spezieller Strukturierung können archiviert werden, wenn die Struktur im Dokument oder in einer separaten Datei erläutert und mitarchiviert wird.

Alle Textdokumente sollten Unicode für die Zeichenkodierung verwenden, wobei UTF-8 ohne BOM besonders empfohlen wird, falls keine speziellen Anforderungen dagegen sprechen. Wenn die verwendete Zeichenmenge es erlaubt, ist ASCII ebenfalls geeignet.

Hinweis: Eingebettete Bilder oder andere Medien sollten zusätzlich separat gespeichert werden. Außerdem muss beachtet werden, dass Links oder dynamische Inhalte, nicht immer dauerhaft erhalten bleiben.

Format	Begründung
 PDF/A	Wenn neben dem Inhalt auch das Aussehen des Dokumentes erhalten bleiben soll und die Bearbeitung des Dokuments abgeschlossen ist, eignet sich PDF/A am besten. Nähere Informationen sind in dem Abschnitt über PDF-Dokumente ab Seite 17 zu finden.
ODT	ODT basiert auf XML und ist Teil vom Open-Document Format. Damit können bearbeitbare Dokumente mit Formatierungsangaben gespeichert werden. ODF verwendet standardmäßig UTF-8 und erlaubt das Einbetten von Fonts.
DOCX	DOCX ist das auf XML basierende Format von Microsoft, das ebenfalls bearbeitbare Dokumente mit Formatierungsangaben speichern kann. DOCX verwendet standardmäßig UTF-8 und erlaubt das Einbetten von TrueType-Fonts.
TXT und <i>plain text</i>	Das Format eignet sich für reinen Text ohne Formatierungsangaben, wie Kursiv-, Fettschreibung oder Schriftgrößen. Die Zeichen sollten in UTF-8 ohne BOM kodiert sein.
strukturierter Text	Alle anderen textbasierten Formate, wie beispielsweise valide SGML-, XML- oder HTML-Dateien können ebenfalls archiviert werden. Für SGML und XML ist zusätzlich die DTD-Datei oder ein XML Schema erforderlich. Anders strukturierte textbasierte Dateien benötigen eine Erläuterung der Struktur innerhalb der Datei oder als zusätzliche separate Datei. Die Zeichen sollten in UTF-8 ohne BOM kodiert sein.

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Format	Begründung
~ RTF	RTF ist ein proprietäres Format von Microsoft für den Datenaustausch, das von vielen Programmen unterstützt wird. Wegen möglichen Kompatibilitätsproblemen sollte DOCX oder ODT bevorzugt werden.
SXW	SXW ist ein Vorgängerformat von ODT, weshalb letzteres auch bevorzugt werden sollte.
✗ DOC	Das DOC-Format von Microsoft eignet sich nicht zur Archivierung, da es proprietär ist und die Inhalte nicht textbasiert gespeichert werden.
PDF	Für die Archivierung wurde speziell das Format PDF/A entwickelt, weshalb dieses verwendet werden sollte.

Dokumentation Metadaten für Textdokumente können in vielen Fällen direkt in das Dokument eingetragen werden. Beispielsweise als Deckblatt oder in dafür vorgesehenen Teilen von strukturierten Dokumenten. Zusätzlich können einige Informationen als Dokumenteigenschaften in der Datei gespeichert werden.

Neben den allgemeinen Angaben zu Einzeldateien, wie sie in dem Abschnitt Metadaten in der Anwendung ab Seite 7 gelistet sind, benötigen Textdokumente insbesondere Angaben zur verwendeten Zeichenkodierung und eine Auflistung der Sprachen.

Falls das Dokument publiziert wurde und eine ISBN oder einen anderen persistenten Identifikator erhalten hat, müssen diese neben den allgemeinen Angaben zur Publikation ebenfalls angegeben werden. Eingebettete Medien, wie Bilder oder Tabellen mit Formeln, sollten separat gespeichert und archiviert werden und in einer Liste weiterer Dateien aufgeführt werden.

Wenn das Aussehen wichtig ist und ein Format verwendet wird, welches das Einbetten von Schriftarten nicht ermöglicht, müssen die verwendeten Schriftarten explizit genannt werden.

Die hier angegebenen Metadaten sind als minimale Angabe zu betrachten und ergänzen die angegebenen Metadaten für Projekte und Einzeldateien in dem Abschnitt Metadaten in der Anwendung ab Seite 7.

Metadatum	Beschreibung
Zeichenkodierung	Welches Zeichenkodierung wird verwendet?
Sprache	In welchen Sprachen ist das Dokument verfasst? Sprachkennungen nach ISO 639 angeben.

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Metadatum	Beschreibung
Identifikator	Wenn das Dokument bereits veröffentlicht wurde und eine ISBN oder einen anderen persistenten Identifikator erhalten hat, sollte dieser angegeben werden.
weitere Dateien	Liste von eingebetteten Medien, die zusätzlich separat gespeichert wurden. Liegt eine Dokumentationsdatei für das Dokument vor, muss diese ebenfalls genannt werden.
Schriftarten	Angabe der verwendeten Schriftarten (Fonts), für Dokumente ohne eingebettete Fonts.

Weitere Metadaten sind methodenabhängig und können in den jeweiligen Abschnitten nachgelesen werden.

Das vollständige Kapitel finden Sie auf:

<https://ianus-fdz.de/it-empfehlungen/textdokumente>

3.3 Bilder – Rastergrafiken

M. Trognitz, P. Grunwald

Bei Rastergrafiken, auch Pixelgrafiken, handelt es sich um digitale Bilder, die mittels rasterförmig angeordneter Bildpunkte, den Pixeln, beschrieben werden. Jedem Pixel ist dabei ein Farbwert zugeordnet. Rastergrafiken haben eine fixe Größe und sind im Gegensatz zu Vektorgrafiken nicht beliebig skalierbar.

Zu den Rastergrafiken gehören: Digitale Fotografien jeder Art, Satellitenbilder, digitalisierte Bilder (Scans), Screenshots sowie digitale Originalbilder und -grafiken.

Langzeitformate Alle als Rastergrafiken vorliegenden Roh- und Urfassungen (Master) von Bildern sind in angemessener Qualität und unkomprimiert im baseline TIFF- oder DNG-Format abzuspeichern. Für georeferenziertes digitales Bildmaterial ist zwecks Erhalt der Referenzdaten das Format GeoTIFF zu verwenden.

Nur für Grafiken, *nicht* für Fotos, eignet sich auch das PNG-Format. Allerdings ist jederzeit das TIFF-Format vorzuziehen. JPEG, bzw. JPG, eignet sich nicht zur Langzeitarchivierung, da es keine verlustfreie Komprimierung anbietet.

Es muss darauf geachtet werden, dass die Bildgröße und -auflösung der originalen Datei erhalten bleibt, wenn in andere Formate umgewandelt wird. Außerdem muss bei der Konvertierung darauf geachtet werden, dass eine verlustfreie Komprimierung verwendet wird. Auch Farbtiefe und Farbraum sollten nach der Konvertierung erhalten bleiben.

Bilder, die Ebenen enthalten, müssen vorher auf eine Ebene reduziert werden. Bei Bedarf, sollten die verschiedenen Ebenen und Komponenten als einzelne Dateien abgespeichert werden.

Hinweis: Wiederholtes Bearbeiten und Abspeichern, führt zu einer allmählichen Abnahme der Qualität. Dieser sogenannte Generationsverlust tritt insbesondere bei der verlustbehafteten Bildkomprimierung auf.

Format	Begründung
✓ Baseline TIFF, unkomprimiert	TIFF ist quasi ein Standardformat für die digitale Langzeitarchivierung von Bilddateien und unterstützt auch die Speicherung von Metadaten im Exif-Format. Obwohl das Format auch Kompression verwenden kann, kommt für die Langzeitarchivierung nur die unkomprimierte Form in Frage.
DNG	Das von Adobe entwickelte Digital Negative Format ist ein offenes Format, das für die Langzeitarchivierung geeignet ist. Damit können RAW-Dateien und deren Metadaten (Exif oder IPTC-NAA) gelesen und gespeichert werden. Außerdem können über XMP weitere Metadaten eingespeist werden.

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Format	Begründung
GeoTIFF	GeoTIFF basiert auf TIFF und ist besonders für georeferenzierte Bilddaten geeignet, da somit die Referenzdaten erhalten bleiben.
~ PNG	PNG ist eine verlustfreie Alternative zu dem GIF-Format, welches eine verlustbehaftete Kompression verwendet. Es bietet eine Farbtiefe von 32 Bit, einen Alphakanal für Transparenz und verlustfreie Kompression. Allerdings wird nur RGB als Farbraum unterstützt und es können keine Exif-Daten gespeichert werden. Das Format eignet sich <i>nicht</i> für digitale Fotos.
✗ JPEG	Trotz einiger Vorzüge, eignet sich das JPEG-Format <i>nicht</i> für die Langzeitarchivierung, da es <i>keine</i> verlustfreie Komprimierung bietet.
GIF	GIF kann sowohl statische, als auch animierte Bilder speichern. Da es aber verlustbehaftet komprimiert, wird PNG als Alternativformat empfohlen.

Dokumentation Eingebettete Metadaten, wie beispielsweise Exif, IPTC-NAA oder XMP, sollten behalten und archiviert werden. Am besten werden sie in eine eigene Text- oder XML-Datei transferiert und getrennt gespeichert.

Neben technischen Informationen, die sich hauptsächlich mit der Erstellung der Rastergrafik befassen, sollten vor allem auch beschreibende und administrative Metadaten über das Bild gespeichert werden.

Hinweis: Werden digitale Aufnahmen mit Programmen bearbeitet, welche die eingebetteten Metadaten ignorieren, gehen diese verloren.

Die hier angegebenen Metadaten sind als minimale Angabe zu betrachten und ergänzen die angegebenen Metadaten für Projekte und Einzeldateien in dem Abschnitt Metadaten in der Anwendung ab Seite 7.

Metadatum	Beschreibung
Identifikator	Name der Datei, z.B. grabung01.tif
Bildunterschrift	Der Titel oder eine passende Bildunterschrift des Bildes
Beschreibung	Beschreibung des Bildes
Urheber	Name des Fotografen oder Erstellers
Datum	Datum der Erstellung oder letzten Änderung des Bildes

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Metadatum	Beschreibung
Rechte	Details zum Urheberrecht
Schlagworte	Schlagworte, wie z.B. Periode, Fundstelle oder charakteristische Merkmale. Wenn vorhanden, angemessene Thesauri verwenden
Ort	Ortsinformationen zu dem Bild. Möglichst in einem standardisierten Format angeben, wie z.B. Lat/Long oder Schlagworte aus einem geeigneten Thesaurus, z.B. Getty Thesaurus of Geographic Names oder GeoNames
Dateiformat & Version	z.B. Baseline TIFF 6.0
Dateigröße	Größe der Datei in Bytes
Bildgröße	Maße des Bildes gemessen in Pixeln, z.B. $400px \times 700px$
Auflösung	Bildauflösung, gemessen in Punkten pro Zoll (dpi)
Farbraum	Der in dem Bild verwendete Farbraum, z.B. RGB oder Graustufen
Farbtiefe	z.B. 24 bit oder 8 bit
Aufnahmegerät	Beispielsweise Details zur Kamera oder dem Scanner
Software	Software mit der das Bild aufgenommen, erstellt oder bearbeitet wurde, wie z.B. Adobe Photoshop CS3

Weitere Metadaten sind methodenabhängig und können in den jeweiligen Abschnitten nachgelesen werden.

Das vollständige Kapitel finden Sie auf:

<https://ianus-fdz.de/it-empfehlungen/rastergrafiken>

3.4 Bilder – Vektorgrafiken und CAD-Daten

M. Trognitz, I. Schwarzbach, F. Martin

Vektorgrafiken sind Grafiken, die mittels grafischer Primitiven, wie beispielsweise Linien, Punkte, Polygone, Kreise und Kurven beschrieben werden. Sie sind daher im Gegensatz zu Rastergrafiken ohne Qualitätsverlust beliebig skalierbar. Zeichnungen von Scherben, Isometrische Darstellungen von Objekten, oder aus CAD- oder GIS-Anwendungen exportierte Pläne werden häufig als Vektorgrafik angelegt oder gespeichert.

Auch bei Daten, die mittels CAD entstehen, handelt es sich um Vektordaten mit grafischen Primitiven, die in den entsprechenden Programmen zusätzliche Funktionalitäten bieten, die speziell auf architektonische und technische Zeichnungen zugeschnitten sind. CAD-Daten können sowohl in 2D als auch in 3D vorliegen und werden in den Altertumswissenschaften meist für Befundpläne von Ausgrabungen, archäologische Karten oder architektonische Pläne verwendet.

Dieser Abschnitt beschäftigt sich mit der Archivierung von zweidimensionalen Vektorgrafiken und CAD-Daten. Die Archivierung von GIS-Daten wird in dem eigenen Abschnitt GIS ab Seite 30 beschrieben. Hinweise zur Archivierung von dreidimensionalen Objekten sind in dem Abschnitt 3D und Virtual Reality ab Seite 52 zu finden.

Langzeitformate Für zweidimensionale Vektorgrafiken ist das vom W3C entwickelte und seit 2001 auch empfohlene Format Scalable Vector Graphics (SVG) für die Langzeitarchivierung geeignet, da es ein XML-basiertes, offen dokumentiertes Format ist, eine weite Verbreitung gefunden hat und als MIME-Type `image/svg+xml` registriert ist. Die aktuelle Version ist 1.1. Neben grafischen Primitiven, geometrischen Formen, Text und eingebetteten Rastergrafiken können SVG-Dateien auch Skripte und Animation enthalten. Mit Hilfe des Gruppierungselementes `<g>` können Ebenen ausgezeichnet werden, wobei dieses Element in der Praxis auch für andere Zwecke genutzt wird. Umfassende Metadaten können in einem eigens dafür vorgesehenen Bereich gespeichert werden. Speziell für mobile Geräte gibt es die Variante SVG Tiny 1.2. Das W3C arbeitet seit 2012 an SVG 2, das in absehbarer Zeit die Version 1.1. ablösen soll. SVG-Dateien können auch komprimiert mittels gzip als SVGZ gespeichert werden, wobei dies nicht für die Archivierung zu empfehlen ist.

Computer Graphics Metafile (CGM) ist unter ISO/IEC 8632 standardisiert, wurde 1986 entwickelt und zuletzt 1999 aktualisiert. CGM kann umfangreiche grafische Informationen und geometrische Primitiven speichern, ist inzwischen jedoch für viele Bereiche veraltet, weshalb SVG für die Archivierung verwendet werden sollte. Das W3C hat 2001 die Variante WebCGM veröffentlicht, die zusätzliche Funktionalitäten für die Verwendung im Internet enthält. Die aktuellste und empfohlene Fassung von WebCGM ist Version 2.1.

Speziell für CAD-Daten gibt es kein Format, das uneingeschränkt für die Langzeitarchivierung zu empfehlen ist. Ein Grund dafür ist die immerwährende und schnelle Entwicklung von CAD-Software, die sich auch auf die Ausgabeformate auswirkt und zu einer hohen Heterogenität und mangelnden Interoperabilität der Datenformate führt. Durch die weite Verbreitung des von der Firma Autodesk vertriebenen Programms AutoCAD haben die von Autodesk entwi-

ckelten Dateiformate DXF und DWG eine weite Verbreitung in der Archäologie und Bauforschung gefunden und sich de facto als Standard durchgesetzt.

Das Drawing Interchange Format (DXF, aktuelle Version v.u.28.1.01) wurde von Autodesk entwickelt, um den Austausch von CAD-Dateien zwischen AutoCAD und anderen Programmen zu ermöglichen. Dabei wurde das Format so gestaltet, dass es möglichst alle in einer Datei enthaltenen Informationen speichern kann, ohne dass dafür die Optimierungen des proprietären DWG-Formates von Autodesk enthüllt werden müssen. Neue Versionen von DXF werden fast jedes Jahr veröffentlicht und durch neue Funktionalitäten angereichert. Für die jüngsten DXF-Versionen stellt Autodesk die Spezifikationen frei zur Verfügung, wobei allerdings ab Release 13 (v.u.13.1.01) manche Bereiche ungenügend spezifiziert sind, so dass unter Umständen nicht alle Funktionen in anderen Programmen umgesetzt und unterstützt werden. DXF gibt es sowohl in einer textbasierten ASCII-Variante, als auch als Binärformat, wobei für die Archivierung erstere bevorzugt werden sollte. Maße werden von DXF nicht unterstützt, weshalb die Angabe des Maßstabes und der Zeicheneinheiten erforderlich ist.

DWG (abgeleitet vom englischen *drawing*) ist das native Speicherformat von AutoCAD und weder offen dokumentiert noch standardisiert. Mit jeder neuen Version von AutoCAD wird auch eine neue Version des DWG-Formates veröffentlicht. Da AutoCAD aber eine so weite Verbreitung gefunden hat, wird dieses Format auch von anderen CAD-Programmen unterstützt. Eine frei verfügbare nachkonstruierte Spezifikation wird von der Open Design Alliance zur Verfügung gestellt.

Für die Archivierung von CAD-Dateien sollten bis zur Etablierung von geeigneten Archivformaten neben dem Originalformat auch Versionen der Datei als DXF und als DWG vorgehalten werden. Um eine möglichst hohe Nutzbarkeit zu gewährleisten, sollte eine ältere verbreitete Version der Formate verwendet werden, wie beispielsweise DWG 2010 (AC1024) und DXF 2010 (AC1024).

Für alle Vektorgrafiken und insbesondere für CAD-Daten gilt die zusätzliche Empfehlung, verschiedene Druckansichten in geeigneten Formaten, wie etwa PDF/A, zu speichern. Dadurch bleibt das intendiert Aussehen der Ursprungsgrafik erhalten auch wenn die direkte Bearbeitbarkeit verloren geht. Jedoch können Vektorgrafiken größtenteils auch wieder aus einer PDF/A-Datei extrahiert werden. Nähere Informationen zu dem Format sind im Abschnitt über PDF-Dokumente ab Seite 17 zu finden.

Externe zugehörige Dateien müssen ebenfalls in einem geeigneten Archivformat gespeichert werden.

PostScript (PS) ist eine Sprache zur Beschreibung von Seiten, die von Adobe entwickelt wurde, und einen Grundstein des PDF-Formates bildet. Darauf aufbauend wurde das Format Encapsulated PostScript (EPS) entwickelt, das neben der Beschreibung in PostScript auch eine eingebettete Datei enthält, die eine Vorschau auf den Inhalt der Datei ermöglicht. Die Vorschaudatei wird jedoch nicht einheitlich gespeichert, was zu einer eingeschränkten Interoperabilität von EPS führt. Ferner kann EPS keine Ebenen und Transparenz speichern. Für die Langzeitarchivierung sollte daher SVG oder PDF/A verwendet werden.

Proprietäre Formate für Vektorgrafiken, wie etwa die Formate AI von Adobe Illustrator und INDD von Adobe InDesign, sind nicht für die Langzeitarchivierung geeignet. Dies trifft ebenfalls auf das von Autodesk entwickelte Format DWF zu, da es primär für Visualisierungen gedacht ist und nicht für

den vollständigen Datenaustausch. In der Industrie werden als alternative offene Austauschformate auch IGES und STEP für CAD-Daten und IFC für Gebäudedatenmodellierung verwendet.

Format	Begründung
✓ SVG	Ein XML-basierter offener Standard vom W3C. Für die Archivierung sollte unkomprimiertes SVG 1.1 verwendet werden.
~ WebCGM	Vom W3C 2001 als Variante von CGM veröffentlicht, um die Verwendung im Internet zu ermöglichen. Wenn es verwendet wird, dann in Version 2.1, wobei SVG für die Langzeitarchivierung besser geeignet ist.
DXF	Das Drawing Interchange Format wurde von Autodesk entwickelt und ist offen dokumentiert. Da Maße nicht gespeichert werden, ist die Angabe des Maßstabes und der Zeicheneinheiten erforderlich. Um eine hohe Nachnutzbarkeit zu erlauben, sollte eine ältere Version, wie etwa DXF 2010 (AC1024), verwendet werden.
DWG	Ist ein proprietäres Format von Autodesk und sollte nur verwendet werden, wenn eine Speicherung in DXF einen erheblichen Funktionsverlust darstellen würde. Um eine hohe Nachnutzbarkeit zu erlauben, sollte eine ältere Version, wie etwa DWG 2010 (AC1024), verwendet werden.
PDF/A	PDF/A ist für die langfristige Speicherung von zweidimensionalen Vektorgrafiken und Druckansichten von CAD-Daten geeignet. Jedoch bleibt nur das Aussehen erhalten. Nähere Informationen zu dem Format sind im Abschnitt über PDF-Dokumente ab Seite 17 zu finden.
✗ CGM	Ist unter ISO/IEC 8632 standardisiert, jedoch veraltet. Stattdessen sollte SVG verwendet werden.
PS, EPS	Encapsulated PostScript beruht auf PostScript, einer Sprache zur Beschreibung von Seiten für den Druck, die von Adobe entwickelt wurden. Für die Archivierung sollte SVG oder PDF/A verwendet werden.
AI, INDD	Proprietäre Formate von Adobe Illustrator und Adobe InDesign sind nicht für die Langzeitarchivierung geeignet.

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Format	Begründung
DWF	Wurde von Autodesk für die Visualisierung von CAD-Daten entwickelt und ist nicht für die Langzeitarchivierung geeignet.

Dokumentation Neben den allgemeinen minimalen Angaben zu Einzeldateien, wie sie in dem Abschnitt Metadaten in der Anwendung ab Seite 7 aufgelistet sind, werden für Vektorgrafiken und insbesondere für CAD-Daten weitere Angaben benötigt.

Neben technischen Informationen sollten vor allem auch beschreibende und administrative Metadaten über die Datei erfasst werden. Dabei können manche Metadaten auch Bestandteil der Grafik selbst sein, wie beispielsweise eine Legende, die über Zeichnungskonventionen für verwendete Strichstärken oder Farben Auskunft gibt.

Zur Dokumentation von CAD-Daten gehören unter anderem die Angabe der Quellen oder Messdaten, die zur Erstellung der Zeichnung herangezogen wurden, die Dokumentation der wesentlichen Arbeitsschritte, die im Verlauf der Erstellung durchlaufen wurden, die Dokumentation der angewendeten Methoden und eine Beschreibung der Abhängigkeitsverhältnisse zwischen den unterschiedlichen Bestandteilen, wie etwa den Ebenen.

Metadatum	Beschreibung
Identifikator	Name der Datei, z.B. grabung01.dxf
Dateiformat & Version	z.B. DXF Release 14
Farbraum	Der in dem Bild verwendete Farbraum, z.B. RGB oder Graustufen
Farbtiefe	z.B. 24 bit oder 8 bit
Bildunterschrift	Der Titel oder eine passende Unterschrift des Bildes
Beschreibung	Beschreibung der Datei
Urheber	Name des Erstellers oder der Bearbeiter
Datum	Datum der Erstellung oder letzten Änderung der Datei
Rechte	Details zum Urheberrecht
Schlagworte	Schlagworte, wie z.B. Periode, Fundstelle oder charakteristische Merkmale. Wenn vorhanden, angemessene Thesauri verwenden

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Metadatum	Beschreibung
Ort	Ortsinformationen zu der Datei. Möglichst in einem standardisierten Format angeben, wie z.B. Lat/Long oder Schlagworte aus einem geeigneten Thesaurus, z.B. Getty Thesaurus of Geographic Names oder GeoNames
Software	Angaben zu Software (inklusive Plug-ins) und Version mit der die Datei erstellt oder bearbeitet wurde, wie z.B. Autodesk AutoCAD 2012
Externe Dateien	Eine Liste aller externen zugehöriger Dateien, wie etwa Messdaten, Rastergrafiken oder eingebundener Datenbanken
Maßstab	Angabe zum verwendeten Maßstab, bzw. was eine Zeicheneinheit darstellt
Koordinatensysteme	Falls vorhanden, verwendetes Koordinatensystem oder Kartenprojektion angeben
Konventionen	Dokumentation der Bedeutung von Farben, Liniestärken, Linientypen, Füllarten etc.
Ebenen	Angabe der Ebenen, deren Inhalt und der Konventionen für deren Benennung und Befüllung
Aufnahmegerät	Beispielsweise Details zum Laserscanner
Messgenauigkeit	Angaben zur Messgenauigkeit des Aufnahmegerätes
Aufnahmemethode	Angabe zur verwendeten Methode zur Datengewinnung

Weitere Metadaten sind methodenabhängig und können in den jeweiligen Abschnitten nachgelesen werden.

Das vollständige Kapitel finden Sie auf:

<https://ianus-fdz.de/it-empfehlungen/vektorgrafiken>

Verweis GIS:

<https://ianus-fdz.de/it-empfehlungen/node/70>

3.5 Tabellen

M. Trognitz

Tabellen werden verwendet, um Informationen strukturiert in Zellen zu speichern, die in Zeilen und Spalten angeordnet sind. Im Gegensatz zu analogen Tabellen bieten digitale Tabellen, die mit entsprechenden Programmen erstellt werden, eine Vielzahl an weiteren Funktionalitäten. Beispielsweise können Zellinhalte nach bestimmten Kriterien sortiert, dynamisch mittels Formeln erzeugt oder Grafiken aus den Daten generiert werden. Um solche und andere Funktionen zu erhalten, erfordert die Speicherung besondere Aufmerksamkeit.

Langzeitformate Für einfache Tabellen ohne interaktive Elemente wie Formeln oder für Tabellenkalkulationen bei denen es ausreicht die Ergebnisse der Formeln zu speichern, wird ein textbasiertes Format mit Trennzeichen für die Archivierung empfohlen.

Dazu eignet sich beispielsweise das CSV-Format, wobei als Trennzeichen ein Komma (,) und als Textbegrenzungszeichen das Anführungszeichen (") verwendet werden sollte, um den Vorgaben von RFC 4180 gerecht zu werden. Andere Trennzeichen und Textbegrenzungszeichen können in begründeten Ausnahmefällen ebenfalls eingesetzt werden und müssen entsprechend dokumentiert werden. Als Zeichenkodierung sollte UTF-8 ohne BOM verwendet werden.

Ein alternatives textbasiertes Format für Tabellen ist das TSV-Format, das als Trennzeichen das Tabulator-Zeichen (U+0009) verwendet. Auch TSV-Dateien sollten UTF-8 ohne BOM als Zeichenkodierung verwenden.

Bei Dateien mit mehr als einem Arbeitsblatt (Tabellenkalkulationen), die als CSV- oder TSV-Datei gespeichert werden sollen, muss jedes Arbeitsblatt gesondert gespeichert werden. Dabei gilt für die Dateinamen der Arbeitsblätter, dass der Name des Arbeitsblattes an den Namen der Tabellenkalkulation, am besten durch ein Unterstrich (_) getrennt, angefügt wird (z. B. `Tabellenname_Blatt1.csv`, `Tabellenname_Blatt2.csv` usw.).

Tabellenkalkulationen, deren zusätzlichen Funktionalitäten erhalten bleiben sollen, werden am besten in einem offenen auf XML basierenden Format gespeichert, wie beispielsweise XLSX oder ODS. Ersteres ist das Standardformat, das in Microsoft Excel seit 2007 verwendet wird und auch von Microsoft entwickelt wurde. Letzteres ist das Format für Tabellenkalkulationen, welches in OpenOffice oder LibreOffice verwendet wird. ODS ist ein Teil vom OpenDocument Format (ODF) und wurde von einem technischen Komitee unter der Leitung der *Organization for the Advancement of Structured Information Standards* (OASIS) entwickelt.

Grafiken, die in Tabellenkalkulationen anhand der Daten erstellt wurden, müssen zusätzlich exportiert und gesondert in einem geeigneten Format gespeichert werden. Dies gilt ebenfalls für eingebettete Bilder oder andere Medien. Passende Formate sind beispielsweise in dem Abschnitt Rastergrafiken ab Seite 23 oder in dem Abschnitt Vektorgrafiken ab Seite 26 zu finden.

Tabellen können auch im XML-Format gespeichert werden. Es gibt eine ganze Reihe an DTDs oder XSDs, die hier als Grundlage dienen können, wie beispielsweise das Schema der TEI, das OASIS Exchange Table Model oder ISO 12083. Auch das Speichern von Tabellen im HTML-Format ist möglich. In jedem Fall muss die zugrundeliegende DTD oder XSD angegeben und ge-

gebenenfalls mit archiviert werden. Die Dateien sollten UTF-8 ohne BOM als Zeichenkodierung verwenden.

Wenn neben den eigentlichen Daten in den einzelnen Zellen auch das Aussehen der Tabelle archiviert werden soll, kann neben einer CSV-, TSV-, XLSX- oder ODS-Datei zusätzlich eine Version der Tabelle im PDF/A-Format gespeichert werden. Bei der Erstellung von Tabellen sollte aber darauf geachtet werden, dass Informationen nicht nur durch Formatierungsangaben, wie beispielsweise die Farbe von Zellen, vermittelt werden, da je nach gewähltem Format die Formatierungsangaben verloren gehen können.

Hinweis: Obwohl Textverarbeitungsprogramme entsprechende Funktionen bieten, sollten Tabellen auch tatsächlich als Tabellen in einem der hier gelisteten Formate gespeichert werden.

Format	Begründung
✓ CSV	Das textbasierte CSV-Format sollte mit einem Komma als Trennzeichen und mit Anführungszeichen als Textbegrenzungszeichen verwendet werden. Ausnahmen müssen dokumentiert werden. Die Zeichen sollten in UTF-8 ohne BOM kodiert sein.
TSV	TSV (MIME-Type text/tab-separated-values) ist ein textbasiertes Format, welches das Tabulator-Zeichen (U+0009) als Trennzeichen verwendet. Die Zeichen sollten in UTF-8 ohne BOM kodiert sein.
ODS	ODS basiert auf XML und ist Teil vom Open-Document Format. Eingebettete Bilder und Medien müssen gesondert gespeichert werden.
XLSX	XLSX ist das auf XML basierende Format von Microsoft. Eingebettete Bilder und Medien müssen gesondert gespeichert werden.
XML, HTML	Tabellen im textbasierten XML- oder HTML-Format können ebenfalls archiviert werden. XML-Dateien benötigen zusätzlich eine DTD-Datei oder das XML Schema. Die Zeichen sollten in UTF-8 ohne BOM kodiert sein.
~ PDF/A	Wenn neben der Daten auch das Aussehen der Tabelle erhalten bleiben soll, eignet sich PDF/A am besten. Zusätzlich sollten die tabellarischen Daten in einem nachnutzbaren Format, wie etwa CSV, gespeichert werden.
SXC	SXC ist ein Vorgängerformat von ODS, weshalb letzteres auch bevorzugt werden sollte.
✗ XLS	Das XLS-Format von Microsoft eignet sich nicht zur Archivierung, da es proprietär ist und die Inhalte nicht textbasiert gespeichert werden.

Dokumentation Neben den allgemeinen minimalen Angaben zu Einzeldateien, wie sie in dem Abschnitt Metadaten in der Anwendung ab Seite 7 gelistet sind, werden für Tabellen und Tabellenkalkulationen weitere Angaben benötigt.

Um die Verständlichkeit einer Tabelle auch für Dritte zu gewährleisten, müssen Name und Zweck der jeweiligen Tabelle und der einzelnen Arbeitsblätter bekannt sein. Jede Spalte benötigt eine Überschrift, und zusätzlich müssen die verwendeten Formatvorgaben, Abkürzungen, Codes, Wertelisten und sonstige Terminologien dokumentiert werden. Um leere Zellen auch explizit als solche zu kennzeichnen, sollte ein vorher festgelegtes Zeichen (z.B. -) eingetragen und dokumentiert werden. Wenn Maßeinheiten nicht direkt aus der Tabelle ersichtlich sind, müssen diese ebenfalls gesondert dokumentiert werden.

Um sicher zu gehen, dass die Tabelle auch vollständig vorliegt, sollten die Anzahl der Spalten, Zeilen und der Arbeitsblätter angegeben werden.

Tabellen in textbasierten Formaten brauchen Angaben zu den verwendeten Trennzeichen, Textbegrenzungszeichen und der Zeichenkodierung.

Für Tabellenkalkulationen müssen bei Bedarf weitere Informationen zu Relationen, Formeln und Makros dokumentiert werden. Eingebettete Medien, wie etwa Bilder, sollten separat gespeichert und archiviert werden und in einer Liste zugehöriger Dateien aufgeführt werden. Dies gilt ebenfalls für Grafiken, die aus den Daten in der Tabellenkalkulation erzeugt wurden.

Metadatum	Beschreibung
Beschreibung der Tabelle oder des Arbeitsblattes	Welchen Zweck verfolgt die Tabelle oder das Arbeitsblatt?
Bezeichnung der Arbeitsblätter	Auflistung der Bezeichnungen der Arbeitsblätter.
Spaltenüberschrift	Jede Spalte einer Tabelle muss einen Namen haben.
Spaltenbeschreibung	Beschreibung und Auflistung der in der jeweiligen Spalte verwendeten Formatvorgaben, Abkürzungen, Codes, Wertelisten, Eingabekonventionen, Fachvokabulare, Zeichen für leere Zellen oder Maßeinheiten.
Anzahl Spalten	Wie viele Spalten enthält die Tabelle?
Anzahl Zeilen	Wie viele Zeilen enthält die Tabelle?
Anzahl Arbeitsblätter	Anzahl der Arbeitsblätter in einer Tabellenkalkulation.
Trennzeichen	Angabe des verwendeten Trennzeichens bei textbasierten Speicherformaten wie CSV.
Textbegrenzungszeichen	Angabe des verwendeten Textbegrenzungszeichens bei textbasierten Speicherformaten wie CSV.

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Metadatum	Beschreibung
Zeichenkodierung	Angabe der verwendeten Zeichenkodierung bei textbasierten Speicherformaten wie CSV oder TSV.
Relationen	Welche Querverweise gibt es innerhalb der Tabellenkalkulation?
Formeln	Welche Formeln werden in der Tabellenkalkulation verwendet?
Makros	Welche Makros gibt es in der Tabellenkalkulation?
Abgeleitete Grafiken	Aus den Daten erzeugte Grafiken müssen zusätzlich separat gespeichert werden und in der Liste zugehöriger Dateien aufgenommen werden.
Sprache	In welchen Sprachen ist das Dokument verfasst? Sprachkennungen nach ISO 639 angeben.
Identifikator	Wenn das Dokument bereits veröffentlicht wurde und eine ISBN oder einen anderen persistenten Identifikator erhalten hat, sollte dieser angegeben werden.
Weitere Dateien	Liste abgeleiteter Grafiken und eingebetteter Medien, wie Bilder, die zusätzlich separat gespeichert wurden. Liegt eine Dokumentationsdatei für das Dokument vor, muss diese ebenfalls genannt werden.

Weitere Metadaten sind methodenabhängig und können in den jeweiligen Abschnitten nachgelesen werden.

Das vollständige Kapitel finden Sie auf:
<https://ianus-fdz.de/it-empfehlungen/tabellen>

3.6 Datenbanken

P. Gerth, F. Schäfer

Übersicht

Datenbanksysteme (DBS) werden zur Strukturierung, Pflege und Verwaltung von Daten verwendet. Sie bestehen aus zwei Komponenten - einerseits aus der Datenbank, welche den logisch zusammengehörigen Datenbestand umfasst, sowie aus dem Datenbankmanagementsystem (DBMS), der Datenbanksoftware, welche u.a. die Zugriffsrechte regelt und Vorkehrungen zur Datensicherheit gewährleistet. Es gibt verschiedene Datenbankmodelle, von denen relationale Datenbanksysteme am häufigsten verwendet und im Rahmen der IT Empfehlungen betrachtet werden. Dabei werden die Daten in Tabellen (Relationen) strukturiert, wobei Verweise von einer Tabelle auf eine andere möglich sind. Näheres zu Tabellen ist in dem Kapitel Tabellen ab Seite 31 zu finden.

Langzeitformate Für relationale Datenbanken wird eine Archivierung in Form eines Exports der Datenbankinhalte, der Strukturinformationen und weiterer Funktionalitäten in textbasierte Formate empfohlen, welche unabhängig vom verwendeten Datenbankmanagementsystem sind. Neben den Daten in den Tabellen müssen die Datenbankstrukturdefinitionen wie Attributdatentypen, Relationen zwischen den Tabellen und Formeln zwingend mit archiviert werden. Unabhängig vom gewählten Archivformat ist es nötig, die grafische Benutzeroberfläche zu dokumentieren, zum Beispiel in Form von Bildschirmfotos oder eines eventuell vorhandenen Benutzerhandbuches. Eine darüberhinaus gehende Erhaltung der Funktionalität ist zumeist nur mit Hilfe der technisch sehr aufwendigen Emulation möglich, bei der das ursprüngliche Datenbanksystem durch ein anderes System mit ähnlicher Funktionsweise ersetzt wird. Unabhängig von den vorgestellten Archivformaten ist der Umgang mit sogenannten Binary Large Objects (BLOBs). Dabei handelt es sich um intern gespeicherte Mediendateien, meistens Fotos, Zeichnungen, z. B. von Funden, oder PDF-Dokumente. Diese Medien müssen in jedem Fall exportiert, auf die dann externen Dateien muss referenziert und je nach Dateityp in für die Archivierung geeigneten Formaten gespeichert werden (siehe entsprechendes Dateiformatkapitel). Grundsätzlich wird von der Verwendung von BLOBs abgeraten und eine externe Speicherung von Medieninhalten empfohlen.

Ein geeignetes Archivformat ist SQL (Structured Query Language), welches seit 1986 bei der ANSI standardisiert ist und seit 1987 ebenfalls als ISO/IEC 9075 zertifiziert ist. Seitdem wurde der SQL-Standard in sieben Revisionen weiter ausgebaut, wobei SQL:2011 die neueste Version ist. Jedoch wird dieser Standard in den verschiedenen Datenbanksystemen wie MySQL, PostgreSQL oder OracleDB unterschiedlich umgesetzt und erweitert. Daher gibt es je nach Datenbanksystem mehrere spezifische Befehle, welche über die Spezifikation des Standards hinausgehen. In Folge dessen sind die verschiedenen DBMS nicht vollständig kompatibel, weswegen eine exportierte Datenbank im SQL-Format nur dann problemlos und ohne zusätzlichen Aufwand importiert werden kann, wenn das ursprüngliche DBMS verwendet wird. Daher kann SQL für die Archivierung nur dann empfohlen werden, wenn ein einheitlicher ANSI-/ISO-Standard, z. B.

SQL:2008, verwendet wird, welcher von DBMS-spezifischen Befehlen bereinigt ist. Der verwendete Standard muss dokumentiert werden. Der Vorteil in der Speicherung in SQL liegt in der gleichzeitigen Erhaltung der kompletten Daten, der Tabellen und ihrer Relationen zueinander, der Attributspezifikationen sowie der kompletten Metadaten. Nach einem Export, erhält man eine singuläre textbasierte Datei, welche sich mit relativ geringem Aufwand wiederherstellen lässt. Die Speicherung einer Datenbank in einem SQL-Skript erfolgt unter Angabe der SQL-Befehle, welche zum Import der Datenbankinhalte benötigt und ggf. für jeden Datensatz wiederholt werden müssen. Die damit einhergehende Redundanz erhöht die Datenmenge. Datenbanken aus geisteswissenschaftlichen Forschungsprojekten erreichen jedoch nur selten eine Größe, bei welcher dieser Nachteil zum Tragen kommt.

XML eignet sich als Archivformat für Datenbanken aus Gründen der Systemunabhängigkeit, Standardisierung, Lesbarkeit und Erweiterbarkeit. Für die Speicherung der Datenbankinhalte kann XML uneingeschränkt empfohlen werden. Die Beschreibung der Datenbankstruktur sollte mithilfe einer entsprechenden XML-Schema-Definition (XSD) wie der Database Markup Language erfolgen. Alternativ kann die Datenbankstruktur im SQL-Format und als Entitäten-Relationen-Diagramm (ERD) gespeichert werden.

Das 2008 initial vom schweizerischen Bundesarchiv entwickelte und im Rahmen des e-ARK-Projekts weiterentwickelte Format SIARD (Software Independent Archiving of Relational Databases) in der Version 2.0 beruht im Wesentlichen auf den beiden ISO-Standards XML gemäß ISO/IEC 15959:2005 und SQL:2008. Es vereinigt damit die Vorteile einer Datenhaltung in XML mit einer Dokumentation der Datenbankstruktur in SQL. Es ist als eCH-Standard offen und wird unter einer freien Lizenz zur Verfügung gestellt.

Ein alternativer bewährter Ansatz zur Archivierung der Tabellen einer Datenbank ist der Export in das CSV-Format, welches besonders langlebig und weit verbreitet ist. Diese Methode ist technisch mit relativ geringem Aufwand verbunden, besitzt aber einen starken Fokus auf die Datenbankinhalte und geht daher mit einem vollständigen Verlust der Funktionalitäten und Datenbankstruktur einher. Infolgedessen ist es zwingend erforderlich, bei der Speicherung im CSV-Format zusätzlich die Datenbankstruktur in Form von strukturierten Diagrammen, wie dem Entitäten-Relationen-Diagramm (ERD) festzuhalten. Selbst dann ist die Datenbankstruktur nur teilweise und sehr aufwendig rekonstruierbar, weswegen das CSV-Format nur bei einer hybriden Lösung mit der Archivierung der Datenbankstruktur in einem geeigneten Format (z. B. SIARD, SQL, XML/DML) bei gleichzeitiger Speicherung der Daten im CSV-Format, empfohlen werden kann. Als Zeichenkodierung sollte UTF-8 verwendet werden, wobei RFC-4180-konform als Trennzeichen ein Komma (,) und als Textbegrenzungszeichen das Anführungszeichen (") verwendet werden sollte. Mehr dazu ist in dem Kapitel Tabellen ab Seite 31 zu finden.

Bei der JavaScript Object Notation (JSON) handelt es sich um ein einfaches textbasiertes Datenformat, welches für einen schnellen Datentransfer im Internet entwickelt wurde und als RFC 7159 und ECMA-404 standardisiert ist. Es besitzt bei der Speicherung von Daten vergleichbare Vorteile wie XML, ist dabei aber wesentlich kompakter. Zur Zeit gibt es allerdings keinen Standard zur Beschreibung von Datenbankstrukturen in JSON, weswegen auch hier auf eine hybride Archivierungslösung mit der Speicherung der Datenbankinhalte in JSON und gleichzeitiger Speicherung der Datenbankstruktur in einem anderen

Format (z. B. SIARD, SQL, XML/DML) zurückgegriffen werden muss.

Die proprietären Dateiformate von Desktop-DBMS, wie Microsoft Access, dBASE oder FileMaker eignen sich nicht für die Archivierung, da die Formatspezifikationen nicht offen vorliegen und verschiedene Versionen oft nicht untereinander kompatibel sind. Ebenso verhält es sich mit den binären Exportformaten serverbasierter, relationaler DBMS, wie Oracles DMP-Format oder das komprimierte BAK-Format von PostgreSQL. Auch diese folgen keinem allgemeingültigen Standard, liegen nur als Binärdatei vor und sind teilweise proprietär. Eine spätere Nachnutzung ist eng an das jeweilige für alle Nutzer kostenpflichtige Produkt sowie spezielle Tools gebunden.

ODB ist ein Container-Dateiformat des Programms, welche in LibreOffice und OpenOffice enthalten ist. ODB ist ein Teil vom OpenDocument Format (ODF) und wurde von einem technischen Komitee unter der Leitung der Organization for the Advancement of Structured Information Standards (OASIS) entwickelt und unter ISO/IEC 26300 standardisiert. Der ZIP-Container enthält verschiedene Dateien, die zum Betrieb der Datenbank erforderlich sind, wie persönliche Konfigurationsdateien, eine binäre Datei des Datenbankmanagementsystems HSQLDB mit den Datenbankinhalten sowie der Datenbankstruktur und die XML-basierten Formulare und Reports. Aufgrund der binären Datenstruktur kann ODB nicht für die Langzeitarchivierung empfohlen werden.

Keines der genannten und empfohlenen Formate unterstützt und enthält alle notwendigen Informationen und signifikanten Eigenschaften, die für die Archivierung und Nachnutzung einer Datenbank relevant sind. Und einige Informationen, etwa ein Handbuch zur Benutzung, müssen ohnehin gesondert gespeichert werden. Daher ist eine hybride Archivierungsstrategie erforderlich, welche verschiedene Dateien mit unterschiedlichen Dateiformaten umfasst. In der folgenden Tabelle werden die Kategorien von Informationen, welche von den für die Archivierung geeigneten Datenbankformate prinzipiell gespeichert werden können, durch Abkürzungen kenntlich gemacht: I = Datenbankinhalt, S = Datenbankstruktur, F = Funktionalitäten, B = Benutzung. Eine zusätzliche Übersicht ist in dem Abschnitt Vertiefung zu finden.

Format	Begründung
✓ SIARD, SIARD2	SIARD ist offen, frei verfügbar und setzt auf die Standards SQL und XML. Der wesentliche Vorteil von SIARD liegt darin, dass sowohl die Datenbankstruktur als auch die Datenbankinhalte gemeinsam beschrieben und archiviert werden. Für die Archivierung sollte Version 2.0 verwendet werden. Speichert: I, S, F.
SQL	Geeignet für die Langzeitarchivierung bei Verwendung eines offiziellen ISO/IEC 9075 Standards, z.B. SQL:2008. Daten, Datenbankstruktur und ein Großteil der Funktionalität bleiben erhalten. Der verwendete Standard muss dokumentiert sein. Speichert: I, S, F.

Fortsetzung auf der nächsten Seite

Format	Begründung
XML	XML bietet sich insbesondere zur Speicherung der Daten einer Datenbank an. Die Datenbankstruktur sollte mit Hilfe eines XSD-Schemas, wie DBML, beschrieben werden. Alternativ ist dies auch mithilfe einer SQL-Datei oder SIARD möglich. Speichert I und teilweise S.
~ CSV	Das CSV Format ist ein langlebiges, textbasiertes und weit verbreitetes Format. Bei dem Export von Datenbankinhalten in das CSV-Format gehen allerdings Informationen zur Datenbankstruktur und zu Funktionalitäten verloren, weshalb es nur bei einer gleichzeitigen Speicherung der Datenbankstruktur in einem anderen Format (z. B. SIARD, SQL, XML/DBML) zu empfehlen ist. Speichert: I.
JSON	JSON bietet sich als Format zur strukturierten Speicherung von Datenbankinhalten an. Da es keine Standards zur Beschreibung der Datenbankstruktur im JSON Format gibt, muss diese gesondert in einem geeigneten Format (z. B. SIARD, SQL, XML/DBML) gespeichert werden. JSON bietet sich insbesondere für die Archivierung von Daten aus NoSQL Datenbanken an. Speichert: I.
MDB, ACCDB	Binäre, proprietäre Formate von Microsoft Access, die nicht zur Archivierung geeignet sind.
✗ FP5, FP7, FMP12	Binäre, proprietäre Formate von FileMaker, die nicht zur Archivierung geeignet sind.
ODB	Containerformat von OpenOffice, welches verschiedene XML-basierte Dateien, wie z.B. Formulare und Abfragen, und die binären Dateien mit den Inhalten und der Struktur der Datenbank enthält. ODB ist nicht zur Archivierung geeignet. Speichert: teilweise B.
DBF	Binäres, proprietäres Format von dBASE, das nicht zur Archivierung geeignet ist.
BAK, DB, DMP	Die verschiedenen, binären Exportformate von relationalen Datenbanksystemen sind für die Archivierung nicht geeignet.

Dokumentation Die Empfehlungen zur Dokumentation von Datenbanken orientieren sich an verschiedenen Standards, darunter SIARD Metadata Schema, Archival Data Description Mark-up Language (ADDML) und Database Markup Language (DBML). Allen Schemata gemein ist eine differenzierte Beschreibung der verschiedenen Ebenen eines Datenbanksystems. Dazu gehören Metadaten zur allgemeinen Beschreibung der Datenbank als auch spezifische Metadaten zur Beschreibung der hierarchisch aufgebauten Ebenen einer Datenbank: den in SQL Datenbanken vorkommenden Schemen, die mehrere Tabellen beinhalten, welche wiederum mit Hilfe von mehreren Attributen definiert werden können.

Weitergehende Dokumente, die zum Verständnis einer Datenbank, ihrer spezifischen Nutzung oder zur Eigenart der Inhalte notwendig sind, wie beispielsweise Benutzerhandbuch, Bildschirmfotos, ERD-Diagramme, Beschreibungen semantischer Konventionen (wie Wertelisten), Anforderungsanalysen oder Rechte- und Rollendokumentation sollten auf alle Fälle mit archiviert werden.

Ergänzend zu den Metadaten, wie sie in dem Abschnitt Metadaten in der Anwendung ab Seite 7 gelistet sind, sollten folgende minimale Metadaten zur nachhaltigen Beschreibung von Datenbanken aufgenommen werden.

Metadatum	Beschreibung
Datenbankname	Ansprache/Name der Datenbank
Beschreibung der Datenbank	Zusammenfassung über Zweck, Bedeutung und Inhalt der Datenbank
Sprache	Liste der Sprachen, in welchen die Daten eingegeben wurden sowie Sprachkennungen nach ISO 639 angeben.
Identifikator	Falls die Datenbank online öffentlich zugänglich ist, sollte ein persistenter Identifikator (z.B. DOI, URI) oder eine eindeutige Adresse (z.B. URL) angegeben werden.
Rechteinhaber	Personen, Institutionen oder Unternehmen, die Rechte an der Datenbankstruktur und/oder den Datenbankinhalten besitzen.
DBMS	Name des Datenbankmanagementsystems mit Versionsnummer mit der die Datenbank betrieben wurde.
Schemata Liste	Liste der Schemata innerhalb der Datenbank sofern vorhanden.
Benutzer	Liste der Benutzer und ihrer Rolle innerhalb der Datenbank.
Rolle	Liste der verschiedenen Rollen und Gruppen mit ihren definierten Zugriffsrechten.

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Metadatum	Beschreibung
Standard	Name und Version verwendeter Standards, etwa zur Definition des Datentyps eines Attributs, zur Dokumentation von Abfragen oder zu dem Datenbankformat, z.B. SQL:2008 oder SIARD 2.0.
Abgeleitete Dateien	Aus den Datenbankinhalten erzeugte Grafiken, Abbildungen, Diagramme müssen zusätzlich separat archiviert und gelistet werden.
Weitere Dateien	Liste weiterer Dokumente, die für das Verständnis und die Dokumentation einer Datenbank notwendig sind, wie beispielsweise ein ERD oder ein Benutzerhandbuch.
Schema Metadatum	Beschreibung
Schema Name	Name des Schemas.
Schema Beschreibung	Bedeutung und Inhalt des Schemas.
Tabellenliste	Liste der Tabellen im Schema.
Funktion	Name und Art der Funktion (Gespeicherte Funktionen, Sichten).
Funktion Beschreibung	Beschreibung der Funktion hinsichtlich Sinn und Zweck.
Funktion Befehl	Befehle der Funktion nach einem einheitlich festgelegten Standard (Syntax Standard).
Tabelle Metadatum	Beschreibung
Tabelle Name	Name der Tabelle.
Tabelle Beschreibung	Bedeutung und Inhalt der Tabelle.
Anzahl Datensätze	Anzahl der Datensätze in der Tabelle.
Attributliste	Liste der Attribute in der Tabelle.
Attribut Metadatum	Beschreibung
Attribut Name	Name des Attributs.
Attribut Beschreibung	Bedeutung und Inhalt des Attributs. Wird das Attribut mit Hilfe von Einheiten, z. B. metrische Maßeinheiten, ausgedrückt, ist die Angabe der Einheit erforderlich.
Attribut Typ	Datentyp des Attributs nach einem einheitlich festgelegtem Standard (Syntax Standard) z. B. „Integer“ nach SQL:2008.
Kontrolliertes Vokabular	Sofern für das Attribut eine Werteliste vorliegt oder ein Thesaurus verwendet wurde, ist dieses zu dokumentieren.

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Metadatum	Beschreibung
Attribut Schlüssel	Sofern es sich um ein Schlüsselattribut handelt, Benennung der Schlüsselart z. B. Primär- oder Fremdschlüssel.
Fremdschlüssel Referenz	Im Falle eines Fremdschlüssels Angabe des referenzierten Attributs mit Angabe der Tabelle und des Schemas.

Das vollständige Kapitel finden Sie auf:

<https://ianus-fdz.de/it-empfehlungen/datenbanken>

3.7 Video

M. Trognitz; Mit Unterstützung von: H. Lewetz

Videos sind in der altertumswissenschaftlichen Forschung vergleichsweise wenig verbreitet und dienen vor allem der Dokumentation von Arbeitsabläufen, Visualisierung von Ergebnissen oder publikumswirksamen Medienauftritten.

Eine digitale Videodatei enthält sowohl visuelle als auch auditive Inhalte. Diese sind jeweils in einem eigenen Format, dem sogenannten Codec, gespeichert und werden in einem Containerformat zusammengeführt. Je nachdem welches Containerformat verwendet wird, können weitere Informationen, wie Metadaten oder Untertitel gespeichert werden. Zu beachten ist, dass es sich bei manchen Containerformaten gleichzeitig um Codecs handeln kann. Die Digitalisierung von analogem Film- und Videomaterial ist nicht Gegenstand dieser Empfehlungen.

Langzeitformate Bei der Auswahl für die Langzeitarchivierung von digitalen Videos muss nicht nur ein passendes Format, sondern auch ein geeigneter Codec gefunden werden. Dabei gelten für die Auswahl des Codecs und des Containerformats die Kriterien, die für die Wahl von Dateiformaten ab Seite 16 beschrieben werden: Es sollte sich um einen weit verbreiteten, möglichst nicht proprietären und offen dokumentierten Standard handeln, der verlustfreie oder gar keine Kompression anwendet. Außerdem sollte darauf geachtet werden, dass das gewählte Format die Speicherung aller relevanten Elemente des Videos, wie beispielsweise Untertitel, unterstützt.

Das Containerformat Matroska (MKV) erfüllt alle Anforderungen an ein Archivformat. Es wird seit 2003 explizit als offenes Containerformat entwickelt, das modernen Ansprüchen genügt und viele verschiedene Codecs enthalten sowie zusätzliche Elemente, wie Untertitel, speichern kann. Das Format basiert auf einer binären Variante von XML, nämlich EBML (Extensible Binary Meta Language), was eine zukünftige flexible Erweiterung erlaubt, jedoch auch sicherstellt, dass ältere Programme weiterhin damit umgehen können. Zusätzlich ist das Format fehlertolerant und kann bis zu einem gewissen Grad auch beschädigte Dateien wiedergeben. Zusammen mit den Codecs FFV1 für Video und FLAC für Audio wird eine Standardisierung bei der IETF angestrebt, weshalb diese Kombination empfohlen wird.

Motion JPEG 2000 (MJ2) ist ein speziell für die Archivierung entwickeltes Containerformat, das unter ISO/IEC 15444-3:2007 zertifiziert und als MIME-Type video/mj2 registriert ist. Es verwendet einen eigenen Codec, JPEG 2000, der jedes Einzelbild verlustfrei mit JPEG 2000 komprimiert. Zugehörige auditive Inhalte werden von Motion JPEG 2000 ebenfalls unterstützt. Das Format ist rechenintensiv und nicht vollständig lizenzfrei, weshalb es besser sein kann den JPEG 2000 Codec in einem anderen Containerformat, wie beispielsweise MXF zu verwenden. Außerdem ist zu beachten, dass MJ2 auch verlustbehaftete Varianten kennt, die für die Archivierung nicht verwendet werden sollten. Motion JPEG 2000 darf nicht mit dem MPEG-Containerformat oder dem verlustbehafteten Motion JPEG (M-JPEG) verwechselt werden.

Die Familie der MPEG-Containerformate wird von der Moving Picture Experts Group (MPEG) entwickelt und ISO/IEC zertifiziert. Es gibt mehrere Generationen von Standards, darunter MPEG-1 (MPG), MPEG-2 (MPG) oder

MPEG-4 (MP4), die entsprechend unterschiedliche Eigenschaften aufweisen. Die Standards beschreiben nicht nur Containerformate, sondern auch Codecs für Video und Audio. MPEG-1 ist seit 1991 unter ISO/IEC 11172 zertifiziert und beschreibt einen Standard für verlustbehaftete Komprimierung von Audio- und Videodaten, der eine gute Qualität und weite Verbreitung hat. MPEG-1 ist insbesondere für das Audioformat MP3 bekannt, eignet sich aber nicht für die Langzeitarchivierung. Der ebenfalls verlustbehaftete Nachfolger, MPEG-2, ist unter ISO/IEC 13818 zertifiziert und hat eine weite Verbreitung in dem Containerformat MPEG-TS (TS) für Rundfunkübertragungen oder auf DVDs gefunden. MPEG-2 kann nur als Langzeitformat für Dateien verwendet werden, die ursprünglich in diesem Format entstanden sind und nicht mehr bearbeitet werden.

Der unter ISO/IEC 14496 zertifizierte MPEG-4-Standard weist eine höhere Effizienz in der Komprimierung und eine höhere Qualität der Videodaten als seine Vorgänger auf. Er verwendet den Codec H.264/MPEG-4 AVC, der meistens verlustbehaftet verwendet wird, jedoch auch eine kaum verwendete verlustfreie Variante hat. Dieser Codec bietet eine hohe Bildqualität, weshalb er eine weite Verbreitung auf Blu-ray Discs, als Aufnahmeformat oder für das Streaming über das Internet gefunden hat. Teil 14 der MPEG-4-Spezifikation definiert das Containerformat MP4, das auch als MIME-Type video/mp4 registriert ist. Wird dieses Format mit dem H.264-Codec verwendet, kann es zur Langzeitarchivierung verwendet werden, wenn dies entweder dem Ursprungsformat der Videodatei entspricht oder verlustfreie Kompression verwendet wird.

Das Containerformat Material eXchange Format (MXF) ist ein offenes Format, das für den Austausch von auditiven und visuellen Inhalten entwickelt wurde und durch eine Reihe von Standards der Society of Motion Picture and Television Engineers (SMPTE) beschrieben wird. MXF wird vor allem im Bereich Kino und Fernsehen verwendet, wo jeweils spezialisierte Varianten verwendet werden, die nicht immer vollständig untereinander kompatibel sind. MXF ist als MIME-Type application/mxf registriert. Zusammen mit der verlustfreien Variante des Codecs JPEG 2000 stellt MXF eine für die Langzeitarchivierung geeignete Variante dar.

Es gibt eine große Zahl an weiteren Containerformaten, wie etwa MOV und ASF/WMV. Sie eignen sich aber nicht für die langfristige Videoarchivierung, da es sich dabei um proprietäre Formate handelt. Bei Audio Video Interleave (AVI) handelt es sich ebenfalls um ein proprietäres Format, es wird jedoch von einigen Archiven als Archivformat verwendet, da es sich um ein einfaches und robustes Format mit einer großen Verbreitung handelt. Die Formate Ogg und Flash wurden vor allem für das Streaming von Videos entwickelt und sind aufgrund der verwendeten verlustbehafteten Codecs nicht als Langzeitformate geeignet.

Die Wahl des Containerformates hängt von dem zu verwendenden Codec ab, da nicht jeder Codec in jedem Container verwendet werden kann. Bekannte und verbreitete Codecs werden nachfolgend beschrieben.

Der Nachteil bei unkomprimierten Videodateien ist, dass sie sehr viel Speicherplatz (teilweise mehrere GB pro Filmminute) beanspruchen und auch für die Wiedergabe entsprechend leistungsfähige Hardware benötigen. Eine Alternative ist die Verwendung von Codecs mit verlustfreier Kompression, wie etwa FFV1, oder HuffYUV und Lagarith. Der FFmpeg Video Codec 1 (FFV1) wurde im Rahmen des FFmpeg-Projektes entwickelt, ist offen dokumentiert, hat

eine weite Verbreitung gefunden und komprimiert sehr gut und schnell. Diese Eigenschaften machen ihn zu einem empfehlenswerten Codec für die Langzeitarchivierung. HuffYUV und dessen Abspaltung Lagarith wurden eigens für Windows-Systeme entwickelt, weshalb diese nicht für die langfristige Archivierung empfohlen werden können.

Einige Codecs können sowohl verlustfrei als auch verlustbehaftet komprimieren. Dazu gehören: Dirac/Schroedinger, JPEG 2000, H.264/MPEG-4 AVC und H.265/MPEG-H (HEVC). Dirac/Schroedinger wurde beim BBC entwickelt, ist jedoch nicht sehr performant und wird von nur wenigen Programmen unterstützt, weshalb er aktuell nicht für die Archivierung empfohlen werden kann. Bei der Verwendung von JPEG 2000 muss darauf geachtet werden, dass die verlustfrei komprimierende Variante verwendet wird, da dieser Standard auch verlustbehaftet komprimieren kann. H.264/MPEG-4 AVC kann in einer verlustfreien Variante verwendet werden, wenn das entsprechende Programm dies unterstützt. Die verlustbehaftete Variante von H.264/MPEG-4 AVC kann nur für die Archivierung von Dateien verwendet werden, die in diesem Codec entstanden sind und nicht mehr bearbeitet werden. Seit 2013 gibt es den Codec High Efficiency Video Coding (HEVC, auch H.265/MPEG-H), der als Nachfolger für H.264 gedacht ist, jedoch noch nicht vollständig spezifiziert ist. Daher und aus patentrechtlichen Gründen ist dieser Codec nicht für die Archivierung zu empfehlen.

Neben den oben erwähnten Codecs gibt es weitere, jedoch verlustbehaftete Codecs, wie M-JPEG, Theora und DV. Im Gegensatz zu JPEG 2000 ist der Codec Motion JPEG (M-JPEG) nicht für die langfristige Archivierung von Videodaten geeignet, da dieser keine verlustfreie Komprimierung unterstützt. Theora ist der für Ogg-Dateien entwickelte Codec. DV ist ein Codec für digitale Videos auf Videokassetten, der mit der Ablösung von Kassetten durch andere Speichermedien obsolet wird. Theora und DV sind keine geeigneten Codecs für die Langzeitarchivierung.

Hinweise auf geeignete Audioformate und dazugehöriger Metadaten sind in dem Kapitel Audio ab Seite 48 zu finden.

Format	Begründung
✓ Matroska (MKV)	Ein offenes Containerformat, das eine große Bandbreite von Codecs und ergänzenden Inhalten unterstützt. Für die Archivierung können die Codecs FFV1 für Video und FLAC für Audio empfohlen werden. Weitere geeignete Codecs für Matroska sind H.264/MPEG-4 AVC und MPEG-2.
~ Motion JPEG 2000 (MJ2)	Motion JPEG 2000 verwendet den Codec JPEG 2000 und ist unter ISO/IEC 15444-3:2007 zertifiziert.
MP4	Der unter ISO/IEC 14496 zertifizierte MPEG-4-Standard verwendet den Codec H.264/MPEG-4 AVC, der meistens verlustbehaftet verwendet wird, jedoch auch verlustfrei komprimieren kann. Das Containerformat MP4 ist als MIME-Type video/mp4 registriert. Wird dieses Format mit dem H.264-Codec verwendet, kann es zur Langzeitarchivierung verwendet werden, wenn dies entweder dem Ursprungsformat der Videodatei entspricht oder verlustfreie Kompression verwendet wird.
MXF	Das Containerformat Material eXchange Format (MXF) ist ein offenes Format, das durch eine Reihe von Standards der SMPTE beschrieben wird. MXF ist als MIME-Type application/mxf registriert. Zusammen mit der verlustfreien Variante des Codecs JPEG 2000 stellt MXF eine für die Langzeitarchivierung geeignete Variante dar.
MPEG-2	MPEG-2 ist unter ISO/IEC 13818 zertifiziert und komprimiert verlustbehaftet. MPEG-2 kann nur als Langzeitformat für Dateien verwendet werden, die ursprünglich in diesem Format entstanden sind und nicht mehr bearbeitet werden.
MPEG (weitere Varianten)	Mit Ausnahme von MPEG-4 und MPEG-2 können andere MPEG-Varianten nicht für die Langzeitarchivierung empfohlen werden.
✗ AVI	Audio Video Interleave ist ein von Microsoft entwickeltes proprietäres Format mit einer weiten Verbreitung. Es eignet sich nicht für die Langzeitarchivierung.

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Format	Begründung
MOV	Ein verbreitetes proprietäres Format von Apple, das in QuickTime verwendet wird und nicht für die Langzeitarchivierung geeignet ist.
ASF/WMV	Ein proprietäres Format von Microsoft, das nicht für die Langzeitarchivierung geeignet ist.
Ogg	Ein von Xiph entwickeltes und offenes Format für das Streaming von Videos, das jedoch nicht für die Archivierung geeignet ist.
Flash	Ein von Macromedia und dann von Adobe entwickeltes Format für das Streaming von Videos. Es ist nicht für die Langzeitarchivierung geeignet.

Dokumentation Neben den allgemeinen minimalen Angaben zu Einzeldateien, wie sie in dem Abschnitt Metadaten in der Anwendung ab Seite 7 gelistet sind, werden für Videos weitere Angaben benötigt, die insbesondere technische Details dokumentieren.

Die technischen Angaben zu Bildformat, Seitenverhältnis, Bildfrequenz, Bitrate und Codec werden zur korrekten Wiedergabe der Datei benötigt. Angaben zu Länge, Tonkanälen, Profilen und weiteren Inhalten sind zur Prüfung auf Vollständigkeit der Datei erforderlich.

Bereits eingebettete Metadaten, wie beispielsweise Exif, oder Bestandteile im Containerformat sollten behalten und archiviert werden. Am besten werden sie in eine eigene Text- oder XML-Datei transferiert und getrennt gespeichert.

Metadatum	Beschreibung
Länge	Dauer des Videos. Diese Angaben sollten konform zu ISO 8601 erfolgen. Beispiel: P3Y6M4DT12H30M5S (3 Jahre, 6 Monate, 4 Tage, 12 Stunden, 30 Minuten und 5 Sekunden) oder T2H2M (2 Stunden und 2 Minuten)
Beteiligte	Angabe der Beteiligten, wie etwa Autor, Regisseur, Darsteller, Interviewpartner etc.
Bildgröße und Seitenverhältnis	Maße des Bildes gemessen in Pixeln, z.B. $1280px \times 720px$, und Angabe des Seitenverhältnisses für die korrekte Darstellung im Verhältnis Breite zu Höhe, z.B.: 16:9
Bildfrequenz	Angabe der Bildfrequenz in Bildern pro Sekunde, z.B. 25

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Metadatum	Beschreibung
Bitrate Video	Angabe der Datenrate in Bits pro Sekunde, z.B. 863 kbps
Farbraum	Angabe des verwendeten Farbraums, z.B. YUV
Farbtiefe	z.B. 8 bit oder 10 bit
Farbunterabtastung	Angabe der verwendeten Farbunterabtastung, z.B. 4:2:0
Videocodec	Name und Version des verwendeten Videocodecs und welche Kompression verwendet wird
Audiostreams	Anzahl der enthaltenen Audiostreams
Audiocodec	Angaben zu den verwendeten Audiocodecs je enthaltenen Audiostreams
Bitrate Audio	Angabe der Datenrate in Bits pro Sekunde je enthaltenen Audiostreams, z.B. 666 kbps
Abtastrate	Angabe der Abtastrate in Hertz je enthaltenen Audiostreams, z.B. 44.1 kHz
Tonkanäle	Angabe über Anzahl der Tonkanäle je enthaltenen Audiostreams
Profil	Wenn ein bestimmtes Videoprofil für ein Containerformat verwendet wurde, bitte angeben.
Weitere Inhalte	Angabe über weitere Inhalte, die in dem Containerformat enthalten sind oder als zusätzliche Datei vorliegen, wie beispielsweise Untertitel
Aufnahmegerät	Herstellername und Modell des Aufnahmegeräts (z.B. eine Kamera)
Software	Name und Versionsnummer der Software, mit der das Video aufgenommen, erstellt oder bearbeitet wurde, wie z.B. Adobe Premiere Pro CC (2015.1)

Weitere Metadaten sind methodenabhängig und können in den jeweiligen Abschnitten nachgelesen werden.

Das vollständige Kapitel finden Sie auf:

<https://ianus-fdz.de/it-empfehlungen/video>

3.8 Audio

M. Trognitz; Mit Unterstützung von: S. Rohde-Enslin, A. Romeyke

Tonaufnahmen werden in der altertumswissenschaftlichen Forschung eher selten erzeugt. Sie werden für die Aufnahme von Interviews, rekonstruierten Musikinstrumenten im Rahmen der Musikarchäologie oder akustischen Eigenschaften von archäologischen Orten erstellt. Sogar Hörbücher altertumswissenschaftlichen Inhalts existieren mittlerweile.

Eine Audiodatei enthält auditive Inhalte, die durch zusätzliche Komponenten wie Kapiteleinteilungen angereichert sein können. Die zusammenfassende Speicherung aller Inhalte erfolgt in einem Containerformat. Die auditiven Inhalte selbst werden nochmals in einem eigenen Format, dem sogenannten Codec, gespeichert. Je nachdem welches Containerformat verwendet wird, können weitere Informationen, wie Metadaten oder Transkriptionen gespeichert werden.

Langzeitformate Da es sich bei den Dateiformaten für digitale Audiodateien um Containerformate handelt, muss bei der Auswahl für die Langzeitarchivierung nicht nur ein passendes Format, sondern auch ein geeigneter Codec gefunden werden. Dabei gelten für die Auswahl des Codecs und des Containerformats die Kriterien, die für die Wahl von Dateiformaten ab Seite 16 beschrieben werden: Es sollte sich um einen weit verbreiteten, möglichst nicht proprietären und offen dokumentierten Standard handeln, der verlustfreie oder gar keine Kompression anwendet. Außerdem sollte darauf geachtet werden, dass das gewählte Format die Speicherung aller relevanten Elemente der Audiodatei, wie etwa Kapiteleinteilungen, unterstützt.

Da viele Containerformate für Audio nur ein Codec-Format speichern können, werden diese in der Regel unter einem Namen zusammengefasst.

Für die Speicherung von unkomprimierten Audiodaten hat sich lineares PCM (Lineare Puls-Code-Modulation, auch LPCM) als Standard durchgesetzt, das von verschiedenen Containerformaten unterstützt und auch von der IASA (International Association of Sound and Audiovisual Archives) empfohlen wird. Der einzige Nachteil ist der große Speicherplatzbedarf mit etwa 10 MB pro Minute.

Eine verlustfrei komprimierende Alternative für die Langzeitarchivierung stellt der Free Lossless Audio Codec (FLAC) dar, der offen dokumentiert und frei verfügbar ist. FLAC wird beispielsweise von dem Containerformat Matroska unterstützt oder kann als eigenständiges Format verwendet werden.

Das Waveform Audio File Format (WAVE) wurde von Microsoft und IBM als Teil des Resource Interchange File Format (RIFF) entwickelt. Es ist ein offen dokumentiertes aber proprietäres Format, das mehrere Audio-Codecs unterstützt. Aufgrund der weiten Verbreitung wird dieses Format jedoch zusammen mit linearem PCM von der IASA empfohlen und hat sich als Standard de facto durchgesetzt.

Die European Broadcast Union (EBU) hat auf WAVE aufbauend das Format BWF entwickelt, das zusätzlich die Einbettung von Metadaten unterstützt. Für die Langzeitarchivierung sollte dieses Format mit linearem PCM verwendet werden. Für Dateien, die größer als 4 GB sind, hat die EBU das Format RF64/MBWF entwickelt, das zusätzlich die Speicherung von mehr als zwei Tonkanälen erlaubt.

Das Containerformat Matroska (MKA) wird seit 2003 explizit als offenes

Containerformat entwickelt, das modernen Ansprüchen genügt und viele verschiedene Codecs enthalten sowie zusätzliche Elemente, wie Kapiteleinteilungen, speichern kann. Das Format basiert auf einer binären Variante von XML, nämlich EBML (Extensible Binary Meta Language), was eine zukünftige flexible Erweiterung erlaubt, jedoch auch sicher stellt, dass ältere Programme weiterhin damit umgehen können. Zusätzlich ist das Format fehlertolerant und kann bis zu einem gewissen Grad auch beschädigte Dateien wiedergeben. Wenn die Audiodaten in linearem PCM oder FLAC gespeichert werden, kann Matroska für die Langzeitarchivierung empfohlen werden.

Aus der Familie der MPEG-Containerformate, die von der Moving Picture Experts Group (MPEG) entwickelt werden und ISO/IEC zertifiziert sind, stammen das MP3- und das AAC-Format. MP3 (MPEG-1 Audio Layer III) wurde schon in MPEG-1 spezifiziert, das seit 1991 unter ISO/IEC 11172 zertifiziert ist. Es handelt sich dabei um einen Audiocodec mit verlustbehafteter Komprimierung, der jedoch eine weite Verbreitung gefunden hat und in einem gleichnamigen Containerformat gespeichert wird.

Advanced Audio Coding (AAC) ist ein Audiocodec, der als Nachfolger von MP3 im Rahmen von MPEG-2 und MPEG-4 spezifiziert wurde und unter ISO/IEC 13818-7 und 14496-3 standardisiert ist. Auch AAC komprimiert verlustbehaftet und kann in einem gleichnamigen Container gespeichert werden oder beispielsweise auch in dem Containerformat MP4. MP3 und AAC sollten nur dann als Langzeitformat für Dateien verwendet werden, wenn diese ursprünglich in dem Format entstanden sind.

Das von Apple entwickelte Audio Interchange File Format (AIFF) ist nicht für die Langzeitarchivierung geeignet, weil es proprietär ist und hauptsächlich nur auf Apple-Systemen Verbreitung gefunden hat. Das von Microsoft entwickelte Windows Media Audio (WMA) ist ebenfalls ein proprietäres Format, das sich wegen der verwendeten verlustbehafteten Kompression nicht für die Langzeitarchivierung eignet. Das Format Ogg wurde vor allem für das Streaming entwickelt und ist aufgrund der verwendeten verlustbehafteten Codecs Opus und Vorbis nicht als Langzeitformat geeignet.

Format	Begründung
✓ FLAC	Free Lossless Audio Codec ist ein verlustfrei komprimierender Codec, der offen dokumentiert und frei verfügbar ist.
WAVE	Waveform Audio File Format wurde von Microsoft und IBM entwickelt und ist offen dokumentiert aber proprietär. Die Audiodaten sollten als lineares PCM gespeichert werden.
BWF	Das Format BWF baut auf WAVE auf und wurde von der EBU entwickelt, um zusätzlich die Einbettung von Metadaten zu unterstützen. Die Audiodaten sollten als lineares PCM gespeichert werden.
~ Matroska	Ein offenes Containerformat, das eine große Bandbreite von Codecs und ergänzenden Inhalten unterstützt. Die Audiodaten sollten als lineares PCM oder FLAC gespeichert werden.
RF64/MBWF	Wurde von der EBU aus BWF entwickelt, um Dateien zu speichern, die größer als 4 GB sind oder mehr als zwei Tonkanäle beinhalten. Die Audiodaten sollten als lineares PCM gespeichert werden.
AAC/MP4	Advanced Audio Coding ist der Nachfolger von MP3 und unter ISO/IEC 13818-7 und 14496-3 standardisiert. Er kann unter anderem in den Formaten AAC oder MP4 gespeichert werden. Die Daten werden verlustbehaftet komprimiert, weshalb er nur für Dateien verwendet werden darf, die ursprünglich in diesem Format entstanden sind.
MP3	MPEG-1 Audio Layer III ist unter ISO/IEC 11172 zertifiziert und verwendet verlustbehaftete Kompression. MP3 kann nur als Langzeitformat für Dateien verwendet werden, die ursprünglich in diesem Format entstanden sind.
✗ AIFF	Das Audio Interchange File Format (AIFF) von Apple ist nicht für die Langzeitarchivierung geeignet, weil es proprietär ist.
WMA	Windows Media Audio ist ein von Microsoft entwickeltes Format mit verlustbehafteter Kompression.
Ogg	Ein von Xiph entwickeltes und offenes Format, das jedoch nicht für die Archivierung geeignet ist, da die Codecs Opus und Vorbis verlustbehaftet komprimieren.

Dokumentation Neben den allgemeinen minimalen Angaben zu Einzeldateien, wie sie in dem Abschnitt Metadaten in der Anwendung ab Seite 7 gelistet sind, werden für Audiodateien weitere Angaben benötigt, die insbesondere technische Details dokumentieren.

Die technischen Angaben zu Codec, Bitrate, Abtastrate und Abtasttiefe werden zur korrekten Wiedergabe der Datei benötigt und vermitteln einen ersten Eindruck über die Qualität der Datei. Angaben zu Länge, Tonkanälen und weiteren Inhalten sind zur Prüfung auf Vollständigkeit der Datei erforderlich.

Bereits eingebettete Metadaten, wie beispielsweise Exif, oder Bestandteile im Containerformat sollten behalten und archiviert werden. Am besten werden sie in eine eigene Text- oder XML-Datei transferiert und getrennt gespeichert.

Metadatum	Beschreibung
Beteiligte	Angabe der Beteiligten, wie etwa Autor, Komponist, Interpret, Interviewpartner etc.
Länge	Dauer der Audiodatei. Diese Angaben sollten konform zu ISO 8601 erfolgen. Beispiel: P3Y6M4DT12H30M5S (3 Jahre, 6 Monate, 4 Tage, 12 Stunden, 30 Minuten und 5 Sekunden) oder T2H2M (2 Stunden und 2 Minuten)
Audiocodec	Angabe des verwendeten Audiocodecs
Tonkanäle	Angabe der Anzahl der Tonspuren und Benennung des Systems, z.B. 5 (Dolby 5.1)
Bitrate	Angabe der Datenrate in Bits pro Sekunde, z.B. 666 kbps
Abtastrate	Angabe der Abtastrate in Hertz, z.B. 44.1 kHz
Abtasttiefe	Angabe der Anzahl der Quantisierungsstufen als Bittiefe, z.B. 16 bit
Weitere Inhalte	Angabe über weitere Inhalte die in dem Containerformat enthalten sind oder als zusätzliche Datei vorliegen, wie beispielsweise Transkriptionen
Aufnahmegerät	Herstellername und Modell des Aufnahmegeräts (z.B. eines Analog-Digital-Umsetzers oder einer Kamera)
Software	Name und Versionsnummer der Software mit der die Audiodatei aufgenommen, erstellt oder bearbeitet wurde, wie z.B. Audacity 2.1.2

Weitere Metadaten sind methodenabhängig und können in den jeweiligen Abschnitten nachgelesen werden.

Das vollständige Kapitel finden Sie auf:
<https://ianus-fdz.de/it-empfehlungen/audio>

3.9 3D und Virtual Reality

M. Trognitz; Mit Unterstützung von: D. Lengyel, K. Niven, V. Gilissen

Im Gegensatz zu statischen zweidimensionalen Bildern können dreidimensionale Repräsentationen von Objekten aus jeder Richtung betrachtet, skaliert und rotiert werden. Ein Punkt in einem 3D-Modell wird von seiner Lage auf der x-, der y- und der z-Achse eines kartesischen Koordinatensystems beschrieben, wobei die z-Achse in diesem Zusammenhang üblicherweise die Tiefe, seltener die Höhe, angibt.

Virtual Reality (Virtuelle Realität) bezeichnet digitale dreidimensionale Welten, mit welchen in Echtzeit interagiert werden kann.

3D-Inhalte können auf unterschiedliche Weise entstehen: durch manuelle Modellierung, wie Rekonstruktionen von Gebäuden, durch Aufnahme, wie etwa einem 3D-Scan von Objekten, oder durch automatisierte Berechnung aus Fotos, wie etwa Photogrammetrie oder Structure from Motion. Von der Entstehungsweise hängen weitere Angaben für die Dokumentation ab, die über die hier angegebenen hinausgehen. Zusätzlichen Angaben sind in den jeweiligen Abschnitten in dem Kapitel Forschungsmethoden ab Seite 63 zu finden, wobei vor allem die Abschnitte Bauforschung, Geodäsie, Geodatenanalyse und Materialaufnahme von Interesse sind. Außerdem bieten die Ergebnisse des Projektes 3D ICONS umfangreiche Informationen zur Dokumentation von 3D-Aufnahmemethoden und die anschließende Verarbeitung.

Langzeitformate 3D-Inhalte sollten in einem offen dokumentierten, textbasierten Format (ASCII) gespeichert werden. Dies ermöglicht bei Bedarf die Rückentwicklung unabhängig von einem bestimmten Programm.

Das zu verwendende Format hängt von der Entstehungsweise und dem Zweck des 3D-Inhaltes ab, da unterschiedliche Formate unterschiedliche Eigenschaften und Elemente speichern, wie beispielsweise Geometrie, Texturen, Lichtquellen oder Standpunkt und Bildausschnitt (auch Viewport genannt). Eine Übersicht über die Speichereigenschaften von den hier empfohlenen 3D-Formaten wird in Tabelle 3.9 auf Seite 56 gegeben.

Das vom Web3D Consortium entwickelte Format X3D (eXtensible 3D Graphics) ist seit 2006 unter ISO/IEC 19775/19776/19777 zertifiziert und eignet sich sowohl zur Speicherung von einzelnen 3D-Modellen, als auch komplexer 3D-Inhalte, wie etwa Virtual Reality. Es ist das Nachfolgeformat von dem seit 2007 unter ISO/IEC 14772-1 zertifizierten VRML-Format und sollte diesem daher vorgezogen werden.

Ein weiteres für komplexe 3D-Inhalte geeignetes Format ist das von der Khronos Group entwickelte COLLADA (collaborative design activity, DAE), das seit 2012 unter ISO 17506 standardisiert ist.

Das U3D-Format bietet einen ähnlichen Funktionsumfang wie X3D und COLLADA und ist insbesondere für das Teilen von 3D-Inhalten in PDF-Dokumenten gedacht. Es ist nicht für die Langzeitarchivierung geeignet.

Die Formate OBJ, PLY und STL eignen sich nicht für komplexe Szenen mit Lichtquellen oder gar Animationen, bieten aber alle Eigenschaften, um die visuellen Oberflächeneigenschaften eines 3D-Objektes zu speichern.

Aus dem CAD-Bereich stammt das Format DXF, welches neben 2D-Inhalten auch 3D-Inhalte speichern kann. Dieses Format sollte nur verwendet werden,


wenn die 3D-Inhalte mit CAD-Software erstellt wurden. In der Industrie werden als alternative Formate auch IGES und STEP verwendet.

Nicht für die Langzeitarchivierung geeignet sind programmspezifische oder binäre Formate, wie beispielsweise FBX, 3DS, MAX, SKP, BLEND, PRC oder NXS.

Um zukünftigen Nutzern einen schnellen Überblick über die 3D-Inhalte zu bieten, ist die zusätzliche Speicherung von Bild- oder Videodateien, die einen ersten Eindruck des Modells vermitteln, zu empfehlen.

Hinweis: Bei der Konvertierung von einem 3D-Format in ein anderes können bestimmte Informationen verloren gehen, wenn sie von dem Zielformat nicht unterstützt werden. Zusätzlich zu dem gewählten Archivierungsformat sollten die originalen Quelldateien, verwendete Texturdateien, Visualisierungen und alle weiteren Dateien, die für die Benutzung und das Verständnis des 3D-Modells relevant sind aufgehoben werden. In den entsprechenden Abschnitten in dem Kapitel Forschungsmethoden ab Seite 63 sind nähere Details zu finden.

Die Abkürzungen in der folgenden Tabelle geben die unterstützten Eigenschaften der 3D-Formate an: DG = Drahtgittermodell, P = Parametrisch, F = Farbe, X = Textur mittels Bild, B = Bumpmapping, M = Material, V = Viewport und Kamera, L = Lichtquellen, T = Transformationen, G = Gruppierung. Diese Begriffe werden in dem Abschnitt Vertiefung ab Seite 56 erläutert.

Format	Begründung
 X3D	Das X3D-Format wurde vom Web3D Consortium entwickelt und ist seit 2006 ISO-zertifiziert. Es darf nicht mit dem proprietären Format 3DXML verwechselt werden. Speichert: DG, P, F, X, B, M, V, L, T, G und Animationen.
COLLADA	COLLADA (.dae) wurde von der Khronos Group entwickelt und ist seit 2012 ISO-zertifiziert. Es basiert auf XML und speichert: DG, P, B-Rep, F, X, B, M, V, L, T, G und Animationen.
OBJ	Das offene OBJ-Format wurde von Wavefront Technologies entwickelt und hat eine weite Verbreitung gefunden. Material oder Texturen werden in gesonderten MTL- oder JPG-Dateien gespeichert, die auch archiviert werden müssen. Speichert: DG, P, F, X, B, M und G.
PLY	Das Polygon File Format (auch Stanford Triangle Format) ist ein einfaches Format, das an der Universität Stanford für Daten von 3D-Scannern entwickelt wurde. Mittels Erweiterungen könnten weitere Eigenschaften gespeichert werden, die aber nicht von jedem Programm unterstützt werden. Für die Langzeitarchivierung sollte die ASCII-Version verwendet werden. Speichert: DG, F, X, B und M.

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Format	Begründung	
~ VRML	Die Virtual Reality Modeling Language ist das Vorgängerformat von X3D. Die jüngste Version wurde 1997 unter dem Namen VRML97 veröffentlicht. Speichert: DG, P, F, X, B, M, V, L, T, G und Animationen.	
STL	Das Format STL wurde von 3D Systems entwickelt. Es steht für <i>stereolithography</i> oder <i>Standard Tessellation Language</i> und findet weite Verbreitung im Bereich von 3D-Druckern und digitaler Fabrikation. Die ASCII-Variante des STL-Formats kann nur die Geometrie eines 3D-Modells speichern. Die Binärvariante des Formats ist weniger speicherintensiv und kann mit einer entsprechenden Erweiterung auch Farben des 3D-Modells speichern, ist aber nicht für die Langzeitarchivierung geeignet.	
DXF	Das von Autodesk entwickelte Drawing Interchange Format stammt aus dem CAD-Bereich und sollte nur für 3D-Inhalte verwendet werden, die mit CAD-Software erstellt wurden. Speichert: DG, P, CSG, B-Rep, F und G.	
×	U3D	Das Universal 3D Format ist ein 2005 von der Ecma standardisiertes 3D-Format, das vom 3D Industry Forum mit Mitgliedern wie Intel und Adobe Systems entwickelt wurde. Dieses Format ist nur für die Integration von 3D-Modellen in ein 3D-PDF relevant. Speichert: DG, F, X, B, V, L, T, G und Animationen.
FBX	Ein proprietäres Format von Autodesk, für den Datenaustausch mit anderen kommerziellen 3D-Programmen.	
3DS, MAX	Proprietäre binäre Formate von AutoDesk.	
SKP	Natives Format von Google SketchUp.	
BLEND	Natives binäres Format von Blender.	
PRC	PRC (Product Representation Compact) ist ein stark komprimiertes Format für den CAD-Bereich und für die Verwendung in 3D-PDFs, das unter ISO 14739-1 standardisiert ist.	
PDF	3D-PDFs mit eingebetteten Modellen in U3D oder PRC eignen sich für einen unkomplizierten Datenaustausch, jedoch nicht für die Langzeitarchivierung.	

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Format	Begründung
NXS	Nexus ist ein von CNR-ISTI offen entwickeltes Format, für die Web-Visualisierung von 3D-Modellen.

Dokumentation Für die Dokumentation von 3D-Inhalten sollten die Leitsätze der *Londoner Charta* berücksichtigt und eingehalten werden. Dabei ist insbesondere „*Leitsatz 4: Dokumentation*“ wichtig, dessen erster Absatz lautet:

„Es sollen genügend Informationen dokumentiert und weitergegeben werden, um das Verstehen und Bewerten der computergestützten Visualisierungsmethoden und -ergebnisse in Bezug auf die Zusammenhänge und Absichten, für die sie eingesetzt werden, zu ermöglichen“

Dazu gehören unter anderem die Dokumentation der Forschungsquellen, die zur Erstellung des 3D-Modells herangezogen wurden, die Dokumentation der Prozesse, die im Verlauf der Modellerstellung durchlaufen wurden, die Dokumentation der angewendeten Methoden und eine Beschreibung der Abhängigkeitsverhältnisse zwischen den unterschiedlichen Bestandteilen.

Wenn das Dateiformat es erlaubt, sollte ein Teil der Metadaten zusätzlich dort integriert werden. Allerdings kann der Großteil der Metadaten nur extern abgelegt werden.

Die hier angegebenen Metadaten sind als minimale Angabe zu betrachten und ergänzen die angegebenen Metadaten für Projekte und Einzeldateien in dem Abschnitt Metadaten in der Anwendung ab Seite 7.

Metadatum	Beschreibung
Anzahl Eckpunkte	Aus wie vielen Eckpunkten besteht das 3D-Modell?
Anzahl Polygone	Wie viele Polygone hat das 3D-Modell?
Geometrietyp	Welcher Geometrietyp wird verwendet (Drahtgittermodell, parametrisch, CSG, B-Rep etc.)?
Maßstab	Welcher Maßstab liegt vor bzw. was stellt eine Einheit dar?
Koordinatensystem	Wird ein geographisches Koordinatensystem oder ein beliebiges verwendet?
Modellstatus	Handelt es sich bei dem Modell um das ursprünglich erzeugte und unverarbeitete Modell (den Master) oder ist es ein Modell, das mittels weiterverarbeitenden Schritten, wie Füllen von Löchern, Vereinfachung oder Glättung aus dem Master erzeugt wurde?

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Metadatum	Beschreibung
Detaillierungsgrad (LOD) oder Auflösung	Wie detailliert ist das Modell oder welche Auflösung wurde beim 3D-Scan verwendet?
Ebenen	Werden Ebenen verwendet? Wie viele?
Farbe und Textur	Werden Farben oder Texturen verwendet? Wie werden diese gespeichert? Müssen zusätzliche Texturdateien archiviert werden?
Material	Informationen über die Materialeigenschaften des Modells und inwieweit sie physikalisch korrekt sind.
Licht	Informationen über die Lichtquellen und inwieweit sie physikalisch korrekt sind.
Shader	Werden spezielle, erweiterte Shader verwendet?
Animation	Ist das Modell animiert? Welche Art von Animation wird verwendet (Keyframe, motion capture)?
Externe Dateien	Eine Liste aller externen Dateien, die für das Öffnen der Datei notwendig sind (z.B. Texturen)

Weitere Metadaten sind methodenabhängig und können in den jeweiligen Abschnitten nachgelesen werden.

Das vollständige Kapitel finden Sie auf:

<https://ianus-fdz.de/it-empfehlungen/3d>

Verweis 3D und Virtual Reality – Vertiefung:

https://ianus-fdz.de/it-empfehlungen/3d?qt-3d_und_virtual_reality=1

Verweis 3D und Virtual Reality – Vertiefung, Speicherung der verschiedenen Eigenschaften:

https://ianus-fdz.de/it-empfehlungen/3d?qt-3d_und_virtual_reality=1#3d_tab-speichereigenschaften

3.10 Webseiten

D. Hagmann

Eine Webseite stellt eine Ressource aus strukturiertem Text im World Wide Web (WWW) dar und besteht in ihrer einfachsten Form aus einer HTML-Datei. Sie kann via Hyperlinks mit beliebig vielen weiteren Ressourcen vernetzt sein. Webseiten sind ein integraler Bestandteil des WWW im Internet. In der Regel ist eine Webseite Teil einer Website, bzw. eines Webauftrittes, also eines zusammengehörenden Paketes von miteinander vernetzten Webseiten und weiteren Ressourcen.

In der altertumswissenschaftlichen Forschung können Webseiten für die Öffentlichkeit zugängliche Informationen beinhalten, wie etwa Blogbeiträge oder ausführliche Projektbeschreibungen.

Der vorliegende Artikel beschäftigt sich vornehmlich mit der Archivierung einzelner Webseiten und nicht mit der Archivierung ganzer Websitesysteme. Um umfangreiche Websites mit mehreren Webseiten zu archivieren, empfehlen sich Online-Speicherdienste, spezialisierte Internetarchive oder dezidierte Softwarelösungen.

Langzeitformate Webseiten können archiviert werden, wenn die nötigen Nutzungsrechte der Inhalte vorliegen. Die Archivierung kann dabei auf verschiedene Arten erfolgen:

- Speicherung als statisches PDF mit zugehörigen Ressourcen
- lokale Speicherung als HTML oder MHTML mit zugehörigen Ressourcen
- Speicherung bei einem spezialisierten Internetarchiv

Als Optimum ist die Langzeitarchivierung einer Webseite in einer Form, die möglichst wenig Informationsverlust garantiert und einfach umzusetzen ist, anzustreben. Webseiten bestehen zum einen aus mindestens einer strukturierten HTML-Datei, zum anderen aus beliebig vielen via Hyperlinks mit der HTML-Datei verbundenen Ressourcen, die teilweise auf demselben Webserver gespeichert sind, aber auch von jedem anderen Ort im Internet bezogen werden können. Dies stellt jedoch nur den theoretisch sehr einfachen Aufbau dar, praktisch bestehen Webseiten aus einer Vielzahl an weiteren strukturierten Textdateien, die etwa das Design der Website regeln (CSS-Dateien) und können zudem über verschiedenste von anderen Websites bezogene und auf der Webseite eingebettete Inhalte (Videos, 3D-Modelle, interaktive Karten etc.) verfügen. Eine der Hauptintentionen jedweder Webseite ist es verschiedene Informationen nach einem vorgegebenen Design in einer bestimmten Abfolge und einem bestimmten Layout dem Nutzer zu vermitteln, vergleichbar zu einer gedruckten Seite in einem Buch. Bei der Archivierung muss beachtet werden, dass für Webseiten teilweise Dateiformate verwendet werden, die für die Langzeitarchivierung dezidiert ungeeignet sind, etwa JPEGs.

Generell empfiehlt es sich, die der Webseite zugrundeliegenden Daten (z.B. Text und Bilder) als Einzeldateien jeweils separat in einem geeigneten Archivformat zu archivieren. Nähere Informationen zu den Archivierungsformaten sind in den entsprechenden Kapiteln zu finden. Auch werden nicht alle multimedialen

Inhalte, Webanwendungen (z.B. Web-GIS) oder über externe Dienste eingebettete Inhalte mit jeder Archivierungsmethode gespeichert, weshalb in solchen Fällen besonderes Augenmerk auf die Auswahl der zu verwendenden Methode gelegt werden muss.

Eine Webseite kann als PDF mit Hilfe des Webbrowsers und eines PDF-Generators als PDF-Datei gespeichert und anschließend mit entsprechender Software in eine archivierbare PDF/A-Datei konvertiert werden. Informationen dazu finden sich im Abschnitt PDF-Dokumente ab Seite 17. Diese Methode führt praktisch immer zu Änderungen des ursprünglichen Layouts. Jedoch können mittels Plug-ins im Webbrowser oder bestimmten Softwareprogrammen Webseiten unter größtenteiliger Wahrung des Layouts als PDF gespeichert werden. Abschließend muss auch hier eine Konvertierung in das PDF/A-Format vorgenommen werden. Multimediale Inhalte (Videos, 3D-Objekte etc.) werden mit dieser Methode nicht gespeichert.

Die lokale Speicherung einer Webseite aus dem WWW mittels eines Webbrowsers ist einfach möglich und wird durch alle aktuellen Webbrowser unterstützt. Hierbei gilt es jedoch, bestimmte Speicherformate zu beachten, da nicht alle in den Webbrowsern verfügbaren Formate für die Archivierung geeignet sind. Für die Strukturierung und Formatierung von Webseiten werden üblicherweise die Hypertext Markup Language (HTML) oder die Extensible Hypertext Markup Language (XHTML), sowie Cascading Style Sheets (CSS) verwendet. Es handelt sich dabei um Standards, die vom W3C entwickelt und empfohlen werden, weshalb diese in den Versionen HTML5, XHTML5 und CSS 3 für die Archivierung empfohlen werden können.

Es bietet sich hier also die Möglichkeit der Speicherung der Webseite als HTML- oder XHTML-Datei an. HTML-Dateien (und XHTML) archivieren den strukturierten Text und die Hyperlinks, jedoch nicht die verknüpften Ressourcen (z.B. Bilder, multimediale Inhalte oder externe Inhalte), zudem wird hierdurch nicht das Design der Webseite, welches durch CSS geregelt wird, übernommen, da die entsprechenden Dateien nicht gespeichert werden. Um auch die verknüpften und für das Design benötigten Ressourcen zu speichern, können diese automatisch in einen zusätzlichen lokalen Ordner geladen werden. In der Regel handelt es sich dabei um HTML/XHTML- und CSS-Dateien, Grafiken, JavaScript-Dateien sowie gegebenenfalls Java-Applets und Multimedia-Dateien.

Die lokale Speicherung einer Webseite in einer einzigen Datei wird mittels MIME HTML (MHTML) ermöglicht. Es handelt sich um ein textbasiertes Format, das in RFC 2557 spezifiziert wird. In der Regel werden bei MHTML-Dateien das Layout und alle Hyperlinks vollständig übernommen. Auch hier muss das Speichern von eingebetteten Inhalten gegebenenfalls gesondert vorgenommen werden.

Das offen dokumentierte Mozilla Archive Format (MAFF) ermöglicht ebenfalls die Speicherung einer Webseite in Form einer einzelnen Datei. Dabei werden die einzelnen Bestandteile in einem ZIP-Container gespeichert. Da dieses Format aktuell nur von Mozilla Firefox unterstützt wird, sollte für die Archivierung jedoch ein anderes Format vorgezogen werden. Ähnlich verhält es sich mit dem Format Webarchive, das derzeit jedoch nur durch Apples Safari Webbrowser unterstützt wird und daher nicht empfohlen werden kann.

Auch HTML-Dateien mit Data-URIs ermöglichen die Speicherung einer gesamten Webseite meist unter Beibehaltung des Layouts in einer einzigen Datei. Data-URIs ermöglichen es, Ressourcen in HTML einzubetten und sind in RFC


2397 definiert. Es handelt sich dabei um eine spezielle Syntax, mit der binäre Daten als ASCII-Zeichenketten kodiert werden. Da Ressourcen als Data-URIs, wie beispielsweise Bilder, direkt und in menschenunlesbarer Form in die Datei integriert werden, können diese nicht nachgenutzt werden, weshalb von einer Speicherung als HTML mit Data-URIs für die Archivierung abgesehen werden sollte.

Eine andere häufig praktizierte, jedoch eindeutig nicht empfohlene Möglichkeit, stellt die Speicherung von Webseiten in der Form von Screenshots dar. Screenshots werden in der Regel im PNG- oder JPEG-Format gespeichert. Dies hat drei Nachteile: (1) Die Konvertierung erfolgt oft in das JPEG-Format, das zur Archivierung nicht geeignet ist. (2) Die Speicherung als Rastergrafik kann in manchen Fällen aufgrund einer zu niedrigen Auflösung zu Qualitätsverlusten führen. Außerdem wird Text nicht mehr als solcher erkannt und gespeichert. (3) Die Konvertierung der Webseite in eine Grafik führt dazu, dass sämtliche Hyperlinks desintegriert werden.

Es besteht zwar hinsichtlich der Punkte (1) und (2) die Möglichkeit, mit entsprechender Software eine Texterkennung und anschließende Speicherung als PDF/A durchzuführen, jedoch können hinsichtlich Punkt (3) dadurch keine Hyperlinks wiederhergestellt werden.

Ein anderer Ansatz ist die Archivierung einer Webseite über einen spezialisierten Archivierungsdienst. Solche werden etwa durch die Bayerische Staatsbibliothek (mit Anmeldung) oder die Organisation Internet Archive angeboten. Hier werden die Webseiten auf einem Server des Archivierungsdienstes gespeichert und können auf diesen Plattformen wiederum über das WWW abgerufen werden. Diese Dienste sind auch zur Archivierung ganzer Websites geeignet. Für die Archivierung ganzer Websites gibt es das Format Web ARChive (WARC), das seit 2009 als ISO 28500 standardisiert ist und von dem International Internet Preservation Consortium aufbauend auf dem Format ARC entwickelt wurde. In einer WARC-Datei werden alle Seiten, Ressourcen und weitere Komponenten einer Website gespeichert.

Hinweis: Angaben zur Archivierung von Programmen in JavaScript sowie Java (Java-Applets) finden sich im Kapitel Eigene Programme und Skripte, Ausführungen zu multimedialen Inhalten (z.B. 3D-Objekte, Audio oder Video) in den entsprechenden Kapiteln.

Format	Begründung
 PDF/A-1, PDF/A-2	PDF/A ist gezielt als stabiles, offenes und standardisiertes Format für die Langzeitarchivierung unterschiedlicher Ausgangsdateien entwickelt worden.

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Format	Begründung
HTML und XHTML	HTML- und XHTML-Dateien können als streng strukturierte Textdokumente, die vom W3C standardisiert sind, problemlos archiviert werden. Die Datei sollte wohlgeformt und in UTF-8 ohne BOM kodiert sein. Es sollte möglichst HTML5 verwendet werden. Zusätzliche Dateien, wie CSS, JavaScript oder andere strukturierte Textformate müssen ebenfalls archiviert werden. Eingebettete Ressourcen müssen gesondert archiviert werden.
MHTML	MHTML-Dateien können als strukturierte Textdokumente mit genauen Spezifikationen für die Archivierung verwendet werden. Die Archivierung von eingebetteten Inhalten muss gegebenenfalls gesondert erfolgen.
WARC	Web ARChive ist als ISO 28500 standardisiert und dient als Containerformat für mehrere Webseiten einer Website.
~ MAFF	Das Format ermöglicht die Speicherung einer ganzen Webseite samt aller zugehöriger Ressourcen komprimiert und verlustfrei in einem ZIP Container und eignet sich zur Langzeitarchivierung, solange die einzelnen Ressourcen selbst in archivfähigen Formaten vorliegen und Hyperlinks entsprechend aktualisiert werden.
HTML mit Data URIs	HTML-Dateien können als strukturierte Textdokumente, die weit verbreiteten Konventionen folgen und aufgrund der integrierten DTD, die die verwendete Struktur beschreibt, problemlos archiviert werden. Data URIs sind ebenso spezifiziert.
✗ andere PDF-Varianten	Viele gängige PDF-Varianten sind nicht für die Langzeitarchivierung geeignet. Stattdessen sollten entweder die Ausgangsdateien in einem passenden Format archiviert oder eine Migration in ein PDF/A-Format vorgenommen werden.
Screenshots	Screenshots eignen sich nur für die Dokumentation der Optik der Webseite, jedoch nicht für die Archivierung der Inhalte, da diese als Rastergrafik gespeichert werden und so kaum nachnutzbar sind.

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Format	Begründung
Webarchive	Ist ein Format von Apple, das derzeit auch nur von Safari unterstützt wird. Es ist nicht für die Archivierung geeignet.

Dokumentation HTML, XHTML und MHTML verfügen über einen eigenen Dokumentenkopf, in dem verschiedene Metadaten eingebettet werden können. Es sollten Angaben zur verwendeten Zeichenkodierung, dem Titel des Dokumentes, dem/der AutorIn sowie Stichwörter gemacht werden. Ergänzende Metadaten können zusätzlich mit Hilfe eines Kommentars in den Kopfdaten der Datei eingefügt werden. Auch in CSS-Dateien können Metadaten als Kommentar eingetragen werden.

Die hier angegebenen Metadaten sind als minimale Angabe zu betrachten und ergänzen die angegebenen Metadaten für Projekte und Einzeldateien in dem Abschnitt Metadaten in der Anwendung.

Metadatum	Beschreibung
Titel	Titel der Webseite
Kurzbeschreibung	Kurze Beschreibung des Inhaltes.
Stichwörter	Schlagworte, die den Inhalt beschreiben.
Autor	Name des Verfassers oder Erstellers der Datei.
Erstellungsdatum	Datum der Erstellung der Datei, also der Archivierung der Webseite.
Bearbeitungsdatum	Datum der letzten Bearbeitung der Webseite.
Abschaltung Webserver	Datum an dem die Webseite zum letzten Mal online verfügbar war
URI	Internetadresse der archivierten Webseite
Identifikator	Wenn das Dokument bereits veröffentlicht wurde und einen Persistent Identifier erhalten hat, sollte dieser angegeben werden.
Sprache	Angabe der im Dokument verwendeten Sprachen. Sprachkennungen nach ISO639 angeben.
Rechte	Details zum Urheberrecht.
Standard	Name und Version des verwendeten Standards, z.B. HTML5 und CSS 3.
Zeichenkodierung	Angabe der verwendeten Zeichenkodierung, z.B. UTF-8 ohne BOM.

Fortsetzung auf der nächsten Seite

Fortsetzung der vorhergehenden Seite

Metadatum	Beschreibung
Beziehungen	Dateien oder Ressourcen, die mit der Datei zusammenhängen, wozu auch frühere Versionen gehören. Bei der Archivierung einer Website mit mehreren Webseiten müssen die Beziehungen der einzelnen Seiten untereinander dokumentiert werden, beispielsweise mit einer Sitemap.
Versionsnummer	Angabe der Dateiversion, bezogen auf den Inhalt. z.B. 1.3.
Software	Name und Version der für die Archivierung der Seite verwendeten Programme
weitere Dateien	Liste von eingebetteten Medien, die zusätzlich separat gespeichert wurden. Liegt eine Dokumentationsdatei für das Dokument vor, muss diese ebenfalls genannt werden.

Das vollständige Kapitel finden Sie auf:
<https://ianus-fdz.de/it-empfehlungen/webseiten>

4 | Weitere Informationen

M. Trognitz

4.1 Forschungsmethoden

Die Ergebnisse aus verschiedenen in den Altertumswissenschaften angewandten Methoden setzen sich oft aus mehreren verschiedenen Dateien unterschiedlicher Formate zusammen. Dadurch gehen die Ansprüche an die Dokumentation und die erforderlichen Metadaten zu den einzelnen Dateiformaten über die des Kapitels zu Dateiformaten hinaus und unterscheiden sich von Methode zu Methode.

Zu folgenden Methoden finden Sie weitere Informationen auf: <https://ianus-fdz.de/it-empfehlungen/forschungsmethoden>

Geodäsie · Georeferenzierung · Oberflächen- (DOM) und Geländemodellierung (DGM) · Reflectance Transformation Imaging (RTI) · Satellitenmessungen

Die vorhandenen Inhalte decken nur einen kleinen Teil der angewandten Forschungsmethoden in den Archäologien und Altertumswissenschaften ab. Weitere Inhalte sind geplant, wobei dafür auch Autoren gesucht werden. Die folgende Auflistung soll einen Eindruck der möglichen Inhalte vermitteln, wobei weitere Vorschläge und Ergänzungen willkommen sind.

3D-Scanning · Anthropologie · Archäobotanik · Archäometrie · Archäozoologie · Ausgrabung · Bauforschung · Datierungsmethoden · Geoarchäologie · Numismatik · Oberflächenbegehung (Survey) · Textanalyse

Bis wir Ihnen ausführlichere Hinweise zur Verfügung stellen können, können Sie sich auf den Seiten des Archaeology Data Service informieren, wo insbesondere folgende Inhalte zu Forschungsmethoden zur Verfügung stehen.

Dendrochronologie · Fernerkundung · GIS · Geophysik · Laserscanning · Photogrammetrie · UAV Survey · Unterwassersurvey

4.2 Weitere Verweise

Weitere Informationen zur Relevanz und Kuratierung von digitalen Forschungsdaten in den Altertumswissenschaften sind:

- Archaeology Data Service, Guides to Good Practice:
<http://guides.archaeologydataservice.ac.uk/g2gp>
- Verband der Landesarchäologen, Ratgeber-Archivierung:
http://www.landesarchaeologen.de/fileadmin/Dokumente/Dokumente_Kommissionen/Dokumente_Archaeologie-Informationssysteme/Dokumente_AIS_Archivierung/Ratgeber_Archivierung_V1.0.pdf