

Dokumentmanagement, Archivierungs- und Verschlüsselungsverfahren

Dozent: Prof. Dr. Michael Eichberg
Kontakt: michael.eichberg@dhbw-mannheim.de, Raum 149B
Version: 2024-02-16

seit Okt. 2023: Professor an der DHBW Mannheim
4 Jahre: Bundeskriminalamt
12 Jahre: Postdoc am Fachgebiet Softwaretechnik der TU Darmstadt
2007: Promotion zum Dr. Ing. am Fachgebiet Softwaretechnik der TU Darmstadt

Dieser Foliensatz basiert auf Folien von: Klaus Götzer.

Alle Fehler sind meine eigenen.

Dokumenten-Management von *Klaus Götzer, Patrick Maué, und Ulrich Emmert*, dpunkt.verlag, 2023.



Dokumentmanagement, Archivierungs- und Verschlüsselungsverfahren - W3WI_EH304.1

Verortung: 3. Studienjahr (5. oder 6. Semester)

Umfang: 25 Vorlesungsstunden und 50 Stunden Selbststudium (2,5 ECTS)

Übergeordnetes Modul:

Fortgeschrittene Konzepte des Informationsmanagements im Gesundheitswesen (W3WI_EH304)

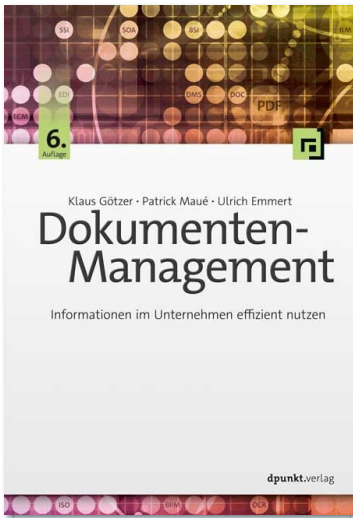
Prüfungsleistung: **Klausur** (60 min/60 Punkte) ~~oder Portfolio~~

Kerninhalte gem. MHB

- Definitionen, Grundsätze, Aufgaben, Funktionen und Prozesse von Dokumentenmanagement
- Archivierungssysteme
- Struktur elektronischer Verwaltungsunterlagen, Archivierungsvarianten der Software- und Hardware-Lösungen
- Rechtliche, technische, organisatorische und wirtschaftliche Anforderungen
- Schnittstellen
- Aktuelle Standards
- Verfügbarkeit elektronischer Verwaltungsakten
- IT-Sicherheit von digital erzeugten und gescannten Dokumenten
- Elektronische Signaturen

Literatur

- Bzgl. Dokumentenmanagement:



- weitere bzgl. der IT sicherheitsrelevanten Aspekte: Verschlüsselungsverfahren und Signaturen

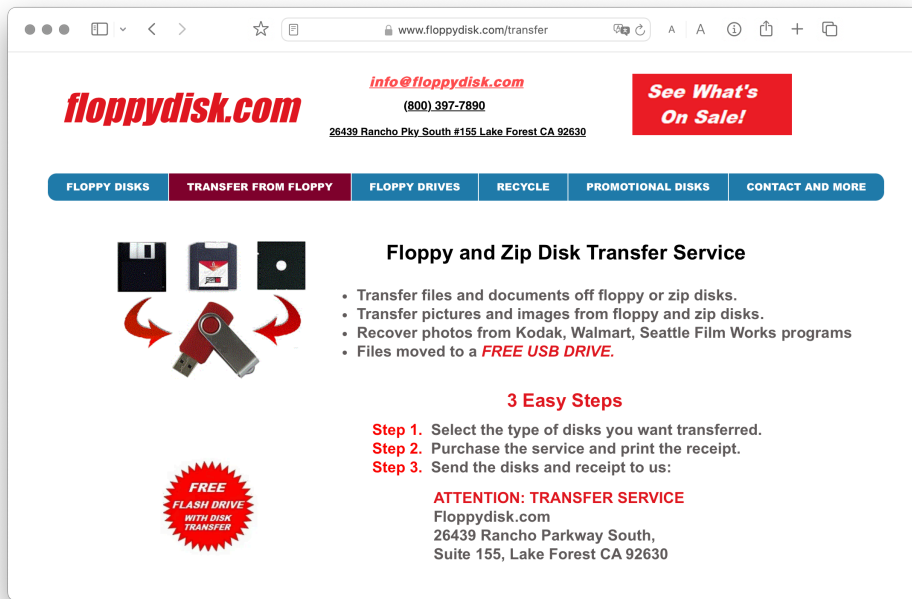
1. WARUM DOKUMENTENMANAGEMENT?

Prof. Dr. Michael Eichberg

Motivation

- Die meisten strukturierten Daten liegen elektronisch vor (ERP, CRM, etc.)
- Dokumente liegen aber (noch immer) in Papierform vor (insbesondere in Deutschland) und es gilt dieses Rationalisierungspotential zu heben
- Alle reden von elektronischen Geschäftsprozessen und Digitalisierung - Dokumente sind ein wesentlicher Bestandteil!
- Wichtige gesetzliche und technische Grundlagen sind geschaffen.
- Experten schätzen, dass 95% der Papierdokumente nach Ablage nicht mehr genutzt werden, da sie zu schwer zu finden sind
- Ähnlich hohe Werte werden für digitale Dokumente genannt, die ohne weitere Strukturierungshilfen in Dateisystemen oder Datenbanken liegen.

Digitalisierung - ein langwieriger Prozess...



Jan. 2024

Digitalisierung - ein langwieriger nicht-aufzuhaltender Prozess...

Japanese government finally bids sayonara to the 3.5" floppy disk

Businesses can at long last submit digital docs to government agencies

Japan is saying sayonara to the floppy disk, which until now was a required medium for submitting some 1,900 official documents to the government.

The announcement (Japanese, machine translated) last week from the Ministry of Economy, Trade and Industry brings decades of physical media submission requirements in Japan to an end. [...]

*Despite being a world leader in cutting-edge technology, Japan has an odd relationship with legacy tech. It's still a land of **cash-only payments and fax machines** that has moved slowly to embrace the modern digital economy.*

—The Register, Mon 29 Jan 2024 // 19:00 UTC

Wie die zahlreichste Bibliothek, wenn ungeordnet, nicht so viel nutzen schafft, als eine sehr mäßige, sorgfältig geordnete; eben so ist die größte Menge von Kenntnissen, die nicht gehörig durch eigenes Denken bearbeitet werden, viel weniger Wert als eine viel geringere Menge von Kenntnissen, die gehörig durchdacht werden.

—Arthur Schopenhauer (mutmaßlich)

Was kann ein Dokumentenmanagement System (DMS) leisten?

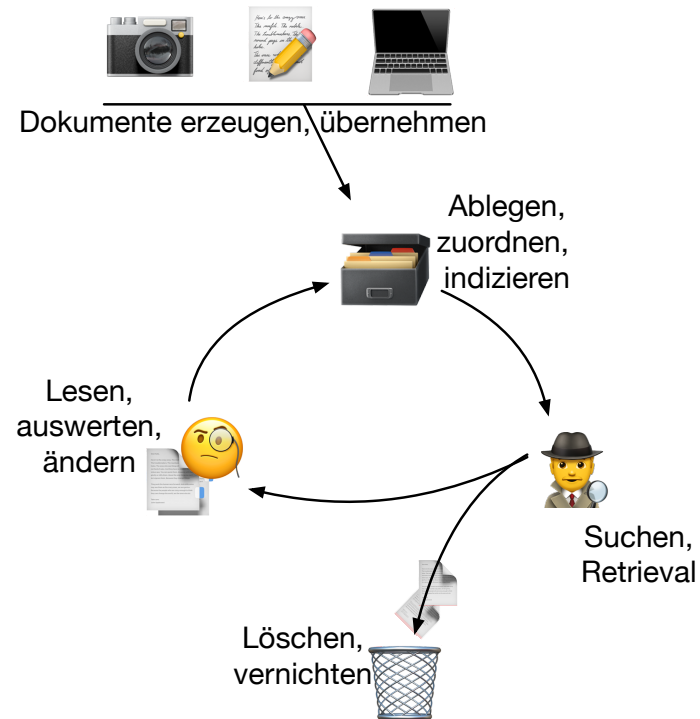
- Beschleunigung der Prozesse
 - Unabhängigkeit des Zugriffs von Ort und Zeit
 - Schnelle Verfügbarkeit der Dokumente
 - Gleichzeitiger Zugriff auf die Dokumente durch mehrere Mitarbeiter
- Revisionssicherheit der Ablage
 - Strukturierte Ablage und Suche von Dokumenten
 - Sicherheit der Dokumente vor Verfälschung und Verlust
 - Transparenz der Prozesse
- Redundanzfreie Archivierung
- Kostenreduktion (Bearbeitungszeiten, Archivkosten....)

Revisionssichere Archivsysteme stellen sicher, dass Informationen wieder auffindbar, nachvollziehbar, unveränderbar und verfälschungssicher archiviert sind.

Gegenüberstellung konventionelles Archiv zu DMS

	Konventionell	DMS
Ablagestruktur	hierarchisch, meist nach Dokumententypen getrennt	datenbankgestützt, Suchbeginn nach jedem Suchkriterium möglich, Suche wahlweise z.B. je Kunde, Vorgang oder Beleg
Auskunftsbereitschaft	zeitverzögert, nach Sichtung/ Entnahme aller Dokumente	unmittelbar durch vorgangsbezogene, bzw. dokumentenübergreifende Suche
Auskunftsaufwand	für Beleg Suche, Entnahme und wieder einordnen, evtl. kopieren und Versand	nur für Recherche am Bildschirm, gegebenenfalls direkter Versand per EMail
Redundanz	Abteilungsablagen = mehrfacher Aufwand	Einmalablage
Vollständigkeit	leidet unter jeder Entnahme	bleibt immer gewahrt
Sicherheit	Ordner u. Dokumente lassen sich leicht entfernen oder kopieren	klare Regelung der Zugriffsrechte, keine ungewollte Entnahme möglich

Zyklus von Dokumenten



Grundlegende Voraussetzungen für Dokumentenmanagementorganisation

In einer Organisation ist zu regeln:

- Kennzeichnung und Beschreibung von Dokumenten
- Fortschreibung und Historienverwaltung von Dokumenten
- Ablage und Archivierung von Dokumenten
- Verteilung und Umlauf von Dokumenten
- Suche nach Dokumenten bzw. Dokumenteninhalten
- Vernichtung von Dokumenten
- Regelung von Verantwortlichkeiten für Inhalt und Verwaltung von Dokumenten

2. DOKUMENT UND DOKUMENTENMANAGEMENT

Prof. Dr. Michael Eichberg

Dokumente - Beispiele



Was ist ein Dokument aus logischer Sicht?

Definition

Ein Dokument fasst inhaltlich zusammengehörende Informationen strukturiert zusammen, die nicht ohne erheblichen Bedeutungsverlust weiter unterteilt werden könnten.

- Die Information ist für einen gewissen Zeitraum zu erhalten.
- Dokumente dienen dem Nachweis von Tatsachen.

Das Dokument ist somit eigentlich der Träger, der die Informationen speichert, egal ob das Dokument ein Stück Papier, eine Datei auf einem Rechner, ein Videoband oder eine Tontafel ist etc.

Was ist ein Dokument aus technischer Sicht?

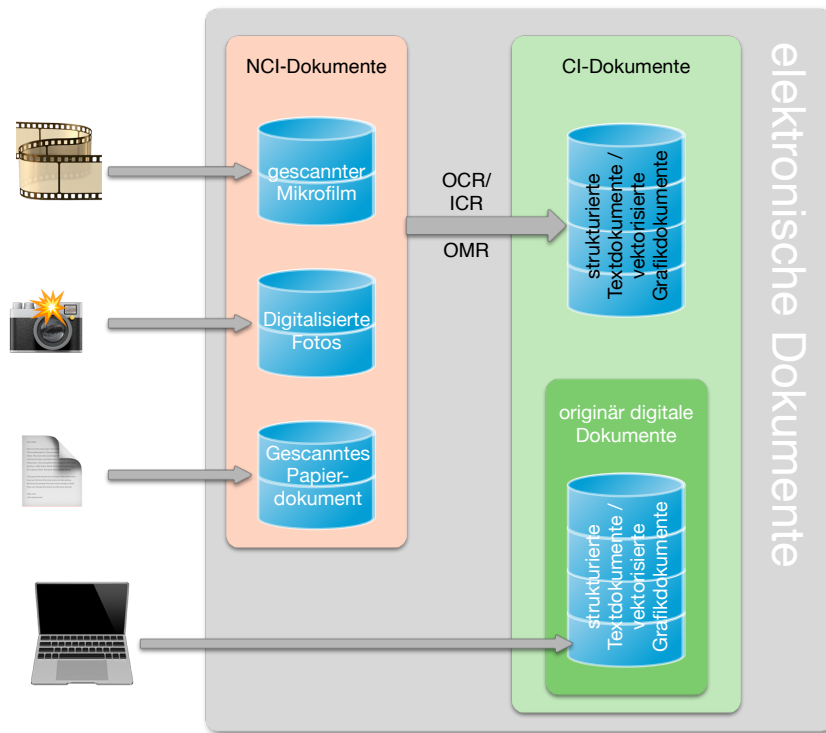
Definition

Ein Dokument ist ein Objekt, das in einer Datenbank beschrieben wird.

Das beschriebene Objekt kann selbst elektronisch gespeichert werden.

Es ist somit eine reine Deklarationsfrage was ein Dokument ist!

Elektronische Dokumente



17

NCI: *Non-Coded-Information (NCI)-Dokumente* sind eingescannte Unterlagen, die als Bild vorliegen, also keine direkte Bearbeitung/Verarbeitung ermöglichen.

CI: Ein *CI-Dokument* ist ein digital erstelltes Dokument, das durch Zeichensätze kodiert ist und von Programmen direkt ausgewertet werden kann.

OCR (Optical Character Recognition):

Text einer gedruckten Vorlage wird durch einfachen Mustervergleich automatisch in maschinenlesbare Zeichen transformiert.

ICR (Intelligent Character Recognition):

Die Qualität der Texterkennung wird durch Kontextanalyse verbessert. Typische Fehler von OCR-Systemen wie zum Beispiel Fehlerkennungen von optisch nahe beieinanderliegenden Zeichen (z.B. "8 und B" oder "0 und O") werden vermieden. (Wie nahe Zeichen beieinander liegen ist stark vom verwendeten Schrifttyp abhängig.)

OMR (Optical Mark Recognition):

liest mit großer Sicherheit spezielle Markierungen in vordefinierten Feldern aus - zum Beispiel in Multiple-Choice-Tests und Vordrucken.

Das "Dokument" →

Eheurkunde

Standesamt Oetzberg

Registernummer Niederschrift über die Eheschließung
E 80/2010

Eheschließung

Ort, Tag Oetzberg, 08.09.2010

Ehemann

Familienname Eichberg

Geburtsname

Vorname(n) Michael

Geburtsdag, Geburtsort

Religion

Familienname nach Eheschließung Eichberg

Geburtsname nach Eheschließung

Ehefrau

Familienname

Geburtsname

Vorname(n)

Geburtsdag, Geburtsort

Religion

Familienname nach Eheschließung

Geburtsname nach Eheschließung

Ort, Tag Oetzberg, 08.09.2010

Urkundsperson *Ohlemüller*
(Ohlemüller, Standesbeamter)

12/1110 Verlag für Standesamtswesen GmbH, Frankfurt am Main · Berlin 2009

→ Eindeutige Dokumenten-ID

→ Stempel

Metadaten des Dokuments:
Erstellungsdatum

Struktur eines Dokumentes



Duale Hochschule Baden-Württemberg - Präsidium
Friedrichstraße 14, 70174 Stuttgart

Herrn
Dr. Michael Eichberg
Birnenweg 21
65205 Wiesbaden

Prof. Dr. Martina Klärle

DHBW
Friedrichstraße 14
70174 Stuttgart
Bearbeiterin:
Frau Krist
Telefon + 49.711.32 06 60-19
Telefax + 49.711.32 06 60-68
lena.krist@dhbw.de
www.dhbw.de
Aktenzeichen
DHBW / 3.2.0.7

Berufung auf eine W2-Professur Wirtschaftsinformatik an der Dualen Hochschule Baden-Württemberg Mannheim,
Verfahrensnummer: 2021-MA-W-4

11.07.2023

Sehr geehrter Herr Dr. Eichberg,

Im November 2022 haben Sie sich auf die genannte Professur beworben. Wir dürfen Ihnen die erfreuliche Mitteilung machen, dass Sie das Präsidium der Dualen Hochschule Baden-Württemberg auf den ersten Listenplatz gesetzt hat.

Wir berufen Sie deshalb auf diese Professur.

Die Einstellung soll Rahmen einer Versetzung und Ernennung zum Professor an der Dualen Hochschule Baden-Württemberg im Beamtenverhältnis auf Lebenszeit erfolgen, ihre Bezüge werden sich nach der Besoldungsgruppe W2 richten. Bitte beachten Sie, dass sich die DHBW für viele Wochen an diesen Ruf gebunden hält. Sie sollten daher innerhalb dieser Frist entscheiden, ob Sie den Ruf annehmen wollen. Aus Gründen der Personalplanung würden wir es jedoch begrüßen, wenn Sie sich möglichst frühzeitig bei uns melden.

Wegen des Einstellungsverfahrens wird sich die DHBW Mannheim mit Ihnen in Verbindung setzen. Für weitere Fragen im Zusammenhang mit dieser Berufung steht Ihnen Herr Prof. Dr. Georg Nagler als Rektor der DHBW Mannheim gerne zur Verfügung (Tel.: 0621 4105 1500; E-Mail: georg.nagler@dhbw-mannheim.de).

Wir würden uns freuen, Sie in Kürze als Professor an unserer Hochschule begrüßen zu dürfen und wünschen Ihnen hierfür bereits viel Erfolg und gutes Gelingen.

Mit freundlichen Grüßen

Prof. Dr. Martina Klärle
Präsidentin

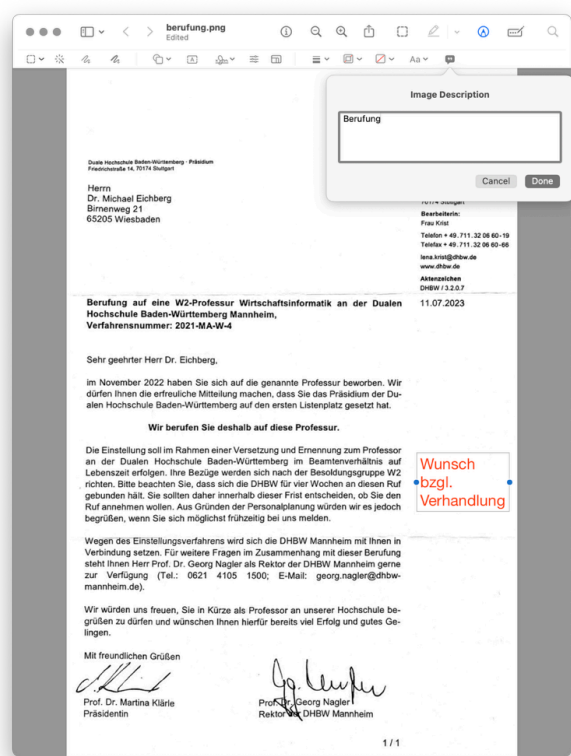
Prof. Dr. Georg Nagler
Rektor der DHBW Mannheim

1 / 1

Kopf

Rumpf

Dokumente annotieren



***Renditions* eines Dokumentes**

- Man unterscheidet zwischen dem Originalformat des Dokuments (z.B. MS-Word oder LibreOffice Format) und Renditions (wie PDF/a und TIFF)
- Formate wie PDF/a und TIFF sind in der Regel langlebiger, es gibt Viewer dafür und erhalten besser den ursprünglichen optischen Zustand.
- Previewimages oft zusätzlich

Hashwerte und Signaturen [1]

- Hashwert ist wie ein mathematischer Fingerabdruck des Dokumentes
- Dieser Hashwert verknüpft mit einer persönlichen Signatur zeigt, dass dieses Dokument von dem Absender der Signatur stammt und das Dokument nicht verändert wurde.

[1] Hashwerte und Signaturen werden wir ausführlich im zweiten Teil der Vorlesung betrachten.

Volltext

- Für Volltextindizierung wird oft der Volltext - insbesondere wenn er durch OCR/ICR gewonnen wurde - mit dem Dokument abgespeichert.
- Die indizierten Begriffe werden oft mit der Angabe der Fundstelle im Dokument abgespeichert, um innerhalb des Dokuments das Suchergebnis anzeigen zu können.

Versionierung von Dokumenten

- Was ist die aktuelle gültige Version?
- Was hat sich gegenüber den Vorgängern geändert?
- Was ist für die nächste in Bearbeitung?
 - Vorgängerversion(en)
 - Freigegebene Version
 - Bearbeitungsversionen
- Versionen des Dokumentes
- Versionen der Metadaten des Dokumentes

Meta-Daten

- Strukturierte Daten, die das Dokument klassifizieren und beschreiben

Beispiele:

- Eindeutige Schlüssel wie Personalnr., Produktnr., ...
- Stichwörter zum Klassifizieren des Textes
- Datum der Erstellung, Änderung, ...
- Autor
- Kategorien wie Mahnung, Anfrage, Branche, Land, ...
- Quelle des Dokuments (Zeitschrift...)

Dies ist insbesondere ein Thema der Datenmodellierung. d.h. welche Daten möchte man wie erfassen.

Arten von Meta-Daten

- Eindeutiger Schlüssel im DMS
- Fremdschlüssel (z.B. Buchungsnummern)
- Statische Metadaten (unveränderlich)
- Dynamische Metadaten (wie Status oder Version der Dokumente)

Beispiel von Meta-Daten einer Verwaltungssoftware für Metadaten

Nr.	Attribut	Muss	Funktion	Quelle	Bemerkung
1	Zeichnungsnummer	M	Eindeutiger Schlüssel	Manuelle Vergabe durch Benutzer	Identifiziert Zeichnung
2	Zeichnungsmappennummer	M	Fremdschlüssel		
3a	Version	M	Version der Zeichnung verwalten	Automatische Vergabe durch DMS Bei Check In	Benutzer entscheidet ob minor oder major
3b	Check-In-Datum	M	Datum des Checkin der Version	Automatische Vergabe durch DMS	Check in Datum
3c	Dokumenten-Owner	M	Gruppe aus letzten Bearbeiter	Aus USER-ID abgeleitet	
3d	Letzter Bearbeiter	M	Identifikation	USER-ID	Beim Check In

Zusammenfassung: Dokumente in einem DMS

Ein Dokument in einem DMS ist ein komplexes Objekt, das aus verschiedenen Komponenten bestehen kann:

- Das Dokument im Originalformat (z.B. odt, docx, xlsx, txt, ...)
- Verschiedene Renditions (pdf, tiff, xml,)
- Vorschaubild
- Volltext
- Annotationen (Layer für Anmerkungen, Stempel, ...)
- Hashwert um elektronische Signaturen zu erzeugen und/oder zu prüfen
- Elektronische Signaturen
- Versionen des Dokumentes
- Metadaten des Dokumentes bzw. der Komponenten des Dokumentes

Dokumentenstrukturen

- Welche Dokumente bilden eine logische Einheit („Mappen“, „Ordner“, „Vorgang“)?
- Metadaten zu diesen Mappen definieren.
- Ein Dokument kann in mehreren Mappen sein.
- Der Inhalt einer Mappe unterteilt sich in:
 1. Dokumente, die immer da sein müssen,
 2. solche die optional da sind und
 3. in nicht vorhersehbare Exoten.

Zusammengesetzte Dokumente (*Compound Documents*)

Komplexes Objekt aus mehreren Dokumenten mit eigener Verwaltungsstruktur:

- Metadaten
- Versionen
- Rechte

3. DOKUMENTENLEBENSZYKLUS

Prof. Dr. Michael Eichberg

Dokumentenlebenszyklus - Überblick



Dokumentenlebenszyklus

Dokumente ...

- entstehen
- verändern sich
- werden festgeschrieben
- dienen als Nachweis / Infoquelle
- müssen bestimmte Zeit aufbewahrt werden
- können bzw. müssen gelöscht werden.

Erstellen von Dokumenten

- Scannen von analogen Dokumenten (Papier, Mikrofilm, ..)
- Neu erstellen von Dokumenten (Vorlagen im DMS,..)
- Vorhandene Dokumente einstellen (drag and drop)
- Dokumente aus Applikationen übernehmen (SAP-Archive-Link, Mail-Archivierung, ...)
- Spezielle Verfahren bei Migration und Massenimporten
- Indizieren der Dokumente entweder automatisch oder manuell

Nutzen und bearbeiten von Dokumenten

- Suchen und Retrieval:
 - Volltext
 - Indizes
 - Verknüpfungen (z.B. in Applikationen)
- Ausgabe der Dokumente auf Bildschirm, Drucker, Mail
- Check out / bearbeiten / Check in

Rahmenbedingungen die Lebensdauer von Dokumenten

- Betrieblichen Notwendigkeiten
- Gesetzlichen Aufbewahrungspflichten
- Datenschutzbestimmungen

- Konzept zur intelligenten Verwaltung, Bewertung und Nutzung von Daten bei möglichst geringen Kosten (Geschäftsregeln, ServiceLevel, ..)
- Betrachtung der Daten und Dokumente über ihren gesamten Lebenslauf aber nicht in Abhängigkeit vom Alter sondern von der Wichtigkeit
- Entwicklung optimaler Verwaltungsstrategien in Abhängigkeit von der aktuellen Wichtigkeit und Nutzung
- Enge Verzahnung von Speicherhardware, Archivierung und Daten-, Dokumenten- und Content-Management.

Löschen von Dokumenten

- Falsche Dokumente (z.B. Fehler beim Indizieren)
- Logisches Löschen
- Physikalisches Löschen

Langfristige Aufbewahrung von Dokumenten

- Technikmuseum (Variante: Alte Umgebung emulieren)
- Dauerhafte Formate nutzen
- Migration der Dokumente auf neue Umgebung

Bewertung von Dateiformaten in Hinblick auf die Dauerhaftigkeit

- vollständige und offene Dokumentation (am Besten mit Standardisierung)
- Plattformunabhängigkeit
- nicht-proprietär (herstellerunabhängig)
- keine "verlustbehaftete" oder proprietäre Komprimierung
- keine eingebetteten Dateien, Programme oder Skripte
- keine vollständige oder teilweise Verschlüsselung
- kein Passwortschutz
- relevante Nutzerbasis

Langfristige Aufbewahrung von Dokumenten

Recommended File Formats for Long-Term Data Curation - Georgia Southern University | University Libraries

Content Type	High probability for long-term preservation	Medium probability for long-term preservation	Low probability for long-term preservation
Text	<ul style="list-style-type: none">■ Plain text (encoding: USASCII, UTF-8, UTF-16 with BOM)■ XML (includes XSD/XSL/XHTML, etc.; with included or accessible schema)■ PDF/A-1 (ISO 19005-1) (*.pdf)	<ul style="list-style-type: none">■ Cascading Style Sheets (*.css)■ DTD (*.dtd)■ Plain text (ISO 8859-1 encoding)■ PDF (*.pdf) (embedded fonts)■ Rich Text Format (*.rtf)■ HTML (include a DOCTYPE declaration)■ SGML (*.sgml)■ Open Office (*.sxw/*.odt)■ OOXML (ISO/IEC DIS 29500) (*.docx)	<ul style="list-style-type: none">■ PDF (*.pdf) (encrypted)■ Microsoft Word (*.doc)■ WordPerfect (*.wpd)■ All other text formats not listed here

Dateiformate und ihre Eignung für die Langzeitarchivierung

SPASSPROJEKT: Entwicklerin erstellt PDF-Dokument in der Größe der Welt

Sind PDFs in ihren Ausmaßen in der Größe limitiert? Eine Frau wollte es genau wissen und erstellte ein Dokument, das größer ist als Deutschland.

PDF-Enthusiastin Alex Chan hat ein Experiment durchgeführt, um ein extrem großes PDF-Dokument zu erstellen – lediglich, um zu sehen, ob es möglich ist. Mit ihrem Wissen über das PDF-Dateiformat machte sich Chan daran, ein PDF-Dokument zu erstellen, das größer ist als die Bundesrepublik Deutschland. [...]

Sie lädt andere Dateiformat-Enthusiasten ein, mit ihr die Möglichkeiten jenseits der dokumentierten Spezifikationen zu erforschen.

—Andreas Donath 3. Februar 2024, 14:21 Uhr

Pflege des Systems

Regelmäßiges Umkopieren der Bestände ist häufig notwendig:

- Datensätze müssen gelöscht werden (Datenschutz)
- Dokumente und Mappen zusammengefasst werden sollen (Performance)
- Datenträger altern (Sicherheit)

Hinweis

- Kann mit einer Migration auf andere Formate bzw. Formatversionen verbunden sein!

4. INTEGRATION

Prof. Dr. Michael Eichberg

DMS als Infrastruktur

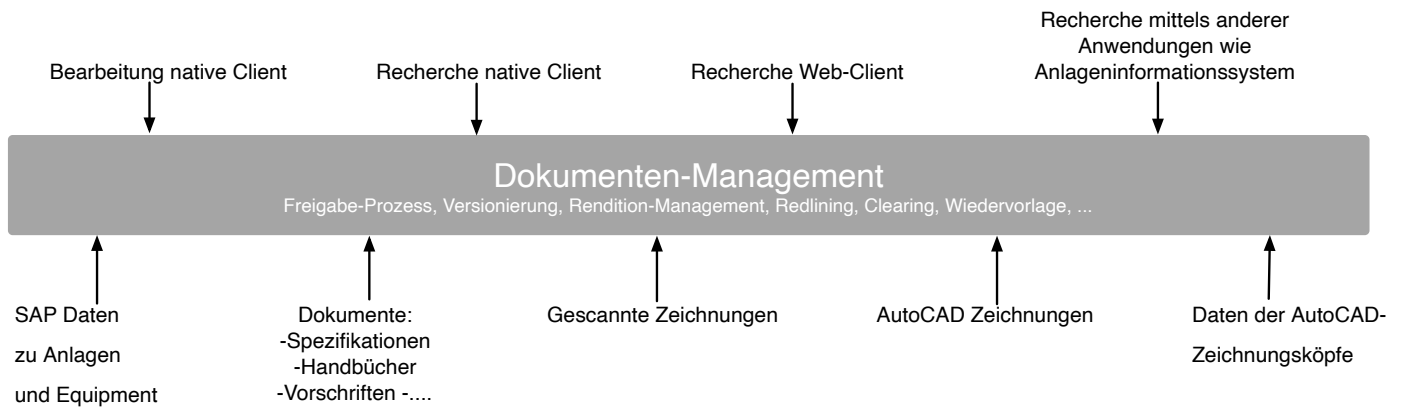
- DMS kann nie sinnvoll für sich allein stehen
- DMS bietet Dienste für andere Applikationen an, um Dokumente zu verwalten.
- Integration in:
 - Standardsysteme (wie MS-Office)
 - Mail-Systeme
 - ERP- / CRM- / PLM-Systeme und andere
 - Suchmaschinen
 - Workflow
 - Scanning
 - ...
- Referenzen auf Dokumente außerhalb des DMS

Typische (Web-) Dienste eines DMS

- Create document
- Move document
- Copy document
- Delete document
- Search document
- Search full text
- Retrieve full document info
- ...

- Create folder
- Move folder
- Copy folder
- Delete folder
- Search folder
- ...

Beispiel: Anlagendokumentation



Redlining bzw. die Redlining-Funktion, oder auch Rotstiftfunktion, bezeichnet das Markieren und Ändern von elektronischen Dokumenten oder Zeichnungen zu Feedback-Zwecken. Anmerkungen und Änderungen sind ersichtlich, ohne dass die Originaldatei verändert wird.

(Vgl. <https://www.fme.de/blog/redlining-spezielle-anwendungsfaelle-auf-basis-von-opentext-documentum-for-life-sciences-teil-5/>)

Beispiel: Anlagendokumentation

- Integrierte und konsistente Gesamtsicht auf alle Anlagen und ihre Bestandteile (Pläne und sonstige Dokumente)
- SAP als führendes System für Schlüssel (Datenqualität)
- Aktuelle Version der Dokumente überall und sofort verfügbar (Web)
- Revisionssichere Prozesse für Bereitstellung und Freigabe der Prozesse
- Eine verbindliche Quelle für alle Dokumente und alle Nutzer
- DMS wird für weitere Bereiche genutzt (z.B. SAP-Eingangs- und Ausgangsrechnungen, Magazin, E-Mail-Archivierung).

Typische Fragestellungen bei Integration

- Welches System ist bezüglich der Schlüssel das führende System
- Einheitliche Nutzer- und Zugriffsrechte
- Wie werden die Systeme synchronisiert (permanent-online oder zyklisch im Batch)
- Schnittstellenrealisierung
- ...