



Thomas Freidl

Konzept und Implementierungsmöglichkeiten von digitalen Bibliotheken an Universitäten

MASTERARBEIT

Technische Universität Graz

Institut für Informationssysteme und Computer Medien (IICM)
Leiter: Univ.-Prof. Dipl.-Ing. Dr.techn. Kappe Frank

Supervisor: Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Scerbakov Nikolai

Graz, September 2012



Thomas Freidl

Concept and implementation of modern digital libraries for universities

MASTER THESIS

Technische Universität Graz

Institut für Informationssysteme und Computer Medien (IICM)
Head: Univ.-Prof. Dipl.-Ing. Dr.techn. Kappe Frank

Supervisor: Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Scerbakov Nikolai

Graz, September 2012

*...für meine Eltern, die immer vorbehaltlos hinter mir
gestanden sind ...*

Kurzfassung

Mit Beginn des digitalen Zeitalters wurden klassische Bibliotheken vor neue Herausforderungen gestellt. Der Einsatz von Computern in nahezu allen Bereichen der Wissenschaft verhalf diesen zu einer noch nie da gewesenen Wissensgenerierung. Es erfolgt dabei eine vollständige Digitalisierung von Informations- und Kommunikationsprozessen. Diese Digitalisierung bietet eine Fülle an neuen Möglichkeiten mit Daten umzugehen, sie zu benutzen, zu verteilen oder zu publizieren. Die Bibliothek hat somit einerseits die Aufgabe diese neuen Technologien anzubieten und zu verwenden, muss aber andererseits die bis dato gewohnte Qualität ihres Angebotes aufrechterhalten. Durch die Verwendung und Bereitstellung dieser digitalen Formate passiert eine Transformation von klassisch gedruckten Medien hin zu digitalen Dokumenten. Dadurch werden Dokumente seltener als Buch und häufiger als Datei angeboten. Durch diesen Übergang wurden zwar klassische Probleme gelöst, aber auch neue geschaffen. Es folgt ein Einblick über Technologien, Formate und Konzepte und deren differenzierte Anwendungsmöglichkeiten. Des weiteren wird eine Beispielimplementierung, eines offenen Anzeigeformates (EPUB), für aktuelle Browser gezeigt.

Suchbegriffe: digitale Bibliothek, Langzeitarchivierung, digitale Dokumente, Ebook Formate, PDF, MOBI, EPUB, DRM, Dublin Core, METS, MARC, MODS, FRBR, Z39.50, SRU/CQL, OAI-PMH, OPDS, Fedora, DSpace, Ebook, EReader

Abstract

With the beginning of the digital era, classic libraries were facing new challenges when an increasing transformation from classic print media to digital documents has been taking place. By means of complete digitization of information and communication processes, the use of computers aided unprecedented knowledge generation, contributing greatly to the distribution and publication of information in many scientific areas. To make new knowledge available to an even broader audience, modern libraries therefore have to make these new technologies accessible to the reader, albeit ensuring high quality of their services. Nowadays, a considerable portion of published information is digital and distributed as a file rather a real paper-based publication. Although classical problems of libraries could be solved with the transition to digitized data, also new questions and problems arose. To elucidate these issues in depth, this thesis provides an overview of current technologies, formats and concepts of digital libraries and different applications thereof. In addition to this review, I will introduce and discuss a sample implementation of an open format (EPUB) for an existing web browser.

Keywords: digital Library, long time preservation, digital documents, e-book formats, PDF, MOBI, EPUB, DRM, Dublin Core, METS, MARC, MODS, FRBR, Z39.50, SRU/CQL, OAI-PMH, OPDS, Fedora, DSpace, EBook, EReader

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Graz, am _____

Datum

Unterschrift

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Graz, _____

Date

Signature

Inhaltsverzeichnis

Abbildungsverzeichnis	xv
Tabellenverzeichnis	xviii
Quellcode	xix
1 Einleitung	1
1.1 Information - von Daten zur Weisheit	2
1.1.1 Daten	2
1.1.2 Information	3
1.1.3 Wissen	3
1.1.4 Weisheit	4
1.2 Konventionelle Bibliothek	5
1.2.1 Aufgaben von konventionellen Bibliotheken	5
1.2.2 Probleme konventioneller Bibliotheken	6
1.2.2.1 Lagerprobleme durch mangelnden Platz für Bücher	6
1.2.2.2 Fehlende Leseräume für Nutzer	6
1.2.2.3 Säurefraß und Schimmelbefall an Büchern	7
1.2.2.4 Distanz von Nutzern	8
1.2.2.5 Entscheidung zwischen Anschaffung und verfügbarem Budget	8
1.2.2.6 Dilemma zwischen Breite des Bedarfs und akutem Bedarf	9
1.2.2.7 Umfangreicher Beschaffungs- und Erschließungsvorgang	9
1.2.2.8 Zeitintensive Fernleihe	9
1.3 Digitale Bibliothek	10
1.3.1 Entstehung neuer Information	10
1.3.2 Aufgaben und Ziele einer digitalen Bibliothek	12
1.3.3 Produkte einer digitalen Bibliothek	12
1.3.4 Probleme digitaler Bibliotheken	13
1.3.5 Langzeitarchivierung	13
1.3.5.1 Bitstream Preservation	13
1.3.5.2 Migration	14

1.3.5.3	Emulation	16
1.3.5.4	Computermuseum	17
1.3.5.5	Mikroverfilmung	19
1.3.6	Domesday Projekt	21
1.4	Vergleich von konventioneller und digitaler Bibliothek	23
1.4.1	Hohe Kosten - Niedrige Kosten	24
1.4.2	Instandhaltungs- und Ersatzprobleme - Wenige Instandhaltungsprobleme	24
1.4.3	Angenehm zu lesen - Weniger angenehm zu lesen	25
2	Grundlagen	26
2.1	Digitale Dokumente	26
2.1.1	Vorteile digitaler Dokumente	26
2.1.1.1	Speicherkapazität	26
2.1.1.2	Ausleihe und Übertragung	27
2.1.1.3	Verfügbarkeit von Exemplaren	27
2.1.1.4	Selektive Verwendung	27
2.1.1.5	Weiterverwendung und Reproduzierbarkeit	28
2.1.1.6	Erweiterte Erschließbarkeit	28
2.1.1.7	Kombination unterschiedlicher Darstellungsmöglichkeiten	28
2.1.2	Nachteile digitaler Dokumente	28
2.1.2.1	Abhängigkeit von Software und Hardware	29
2.1.2.2	Leichte Veränderbarkeit	29
2.1.2.3	Anfällig für Beschädigungen	29
2.1.2.4	Unsicherheit bei Übermittlung	31
2.1.2.5	Problem der Langzeitarchivierung	32
2.2	Dokumentenformate - Voraussetzungen und Anforderungen	33
2.2.1	Offenheit	33
2.2.2	Verbreitung	33
2.2.3	Komplexität	34
2.2.4	Schutzmechanismen	35
2.2.5	Selbstdokumentation	35
2.2.6	Robustheit	35
2.2.7	Abhängigkeit	35
2.3	eBooks	36
2.3.1	Marktentwicklung	37
2.4	Anzeigeformate	37

3	Technologien	41
3.1	Häufig genutzte Anzeigeformate	41
3.1.1	PDF	41
3.1.1.1	Aufbau eines PDF Dokumentes	41
3.1.1.2	Vorteile von PDF [Ran07]	44
3.1.1.3	Nachteile von PDF [Leu09]	45
3.1.2	EPUB	46
3.1.2.1	EPUB Publications 3.0	46
3.1.2.2	EPUB Content Documents 3.0	46
3.1.2.3	EPUB Open Container Format (OCF) 3.0	46
3.1.2.4	EPUB Media Overlays 3.0	47
3.1.2.5	Aufbau eines EPUB Dokumentes	48
3.1.2.6	Beispiel eines EPUB Dokumentes	49
3.1.2.7	Vorteile von EPUB [Pra12]	55
3.1.2.8	Nachteile von EPUB [Pra12]	55
3.1.3	Mobipocket / AZW / KF8	55
3.1.4	Apple / iBooks Author	56
3.2	eBook Bundles	57
3.3	DRM (Digital Rights Management)	58
3.3.1	Hartes DRM	58
3.3.1.1	Grundlegende DRM Architektur für hartes DRM	59
3.3.1.2	Vorteile von hartem DRM	60
3.3.1.3	Nachteile von hartem DRM	61
3.3.2	Weiches DRM	62
3.3.2.1	Vorteile von weichem DRM	65
3.3.2.2	Nachteile von weichem DRM	65
3.3.3	Aktueller DRM Einsatz	66
3.4	Metadaten	67
3.4.1	Deskriptive Metadaten	68
3.4.2	Strukturelle Metadaten	68
3.4.3	Administrative Metadaten	68
3.4.4	Metadatenstandards	69
3.4.4.1	DC (Dublin Core)	69
3.4.4.2	METS (Metadata Encoding & Transmission Standard)	70
3.4.4.3	MARC (Machine Readable Cataloguing)	71
3.4.4.4	MODS (Metadata Object Description Schema)	73
3.4.4.5	FRBR (Functional Requirements for Bibliographic Records)	74
3.5	Austauschprotokolle	76
3.5.1	Z39.50	76
3.5.1.1	SRU/CQL	77

3.5.2	OAI-PMH	77
3.5.3	OPDS	79
4	Konzepte und Systeme	83
4.1	Kriterien bei der Auswahl einer digitalen Bibliothek im Bereich Universität	83
4.2	Fedora	85
4.2.1	Aufbau/Struktur des digitalen Objekts	85
4.2.2	Möglichkeiten der Strukturierung / Relationen	86
4.2.3	Speicherung digitaler Inhalte / Metadaten	87
4.2.4	Suchen und Browsen	87
4.2.5	Rechtmanagement und Authentifizierung	87
4.2.6	Schnittstellen und Interaktionsmöglichkeiten	87
4.2.7	Customizingmöglichkeiten / Erweiterbarkeit	88
4.2.8	Verwendete Technologien und Services	88
4.2.9	Größe der Community	88
4.2.10	Hardware und Softwarevoraussetzungen	88
4.3	DSpace	88
4.3.1	Aufbau/Struktur des digitalen Objekts	89
4.3.2	Möglichkeiten der Strukturierung / Relationen	91
4.3.3	Speicherung digitaler Inhalte / Metadaten	91
4.3.4	Suchen und Browsen	92
4.3.5	Rechtmanagement und Authentifizierung	92
4.3.6	Schnittstellen und Interaktionsmöglichkeiten	92
4.3.7	Customizingmöglichkeiten / Erweiterbarkeit	92
4.3.8	Verwendete Technologien und Services	92
4.3.9	Größe der Community	93
4.3.10	Hardware und Softwarevoraussetzungen	93
4.4	Fazit	93
4.4.1	Weitere Anbieter von Repository Systemen	94
5	Erweiterungen	95
5.1	Online EPUB eBook Reader	95
5.1.1	Erstellung der eReader Oberfläche	96
5.1.2	Parsen der *.ncx Datei	98
5.1.3	Laden und Anzeigen der einzelnen Kapitel	99
5.1.4	Darstellung der einzelnen Seiten	105
5.1.5	Berechnung der Seitenanzahl	106
5.1.6	Anbieten verschiedener EPUB eBooks	108
5.1.7	Wechsel der spezifischen EPUB CSS Dateien	108
5.1.8	Anzeige der Bilder	109
5.1.9	Zweiseiten / Einseiten Layout - online EPUB Reader	111

5.1.10	Browserübergreifende Entwicklung	111
5.1.10.1	Probleme im IE (Internet Explorer)	112
5.1.10.2	Fehlende CSS3 Unterstützung	114
6	Zusammenfassung und Ausblick	115
	Literatur	118

Abbildungsverzeichnis

1.1	Speicherkapazität analoger und digitaler Daten [HLo7]	1
1.2	Prozess: Von Daten zu Wissen nach Nathan Shedroff [She99]	4
1.3	Anzahl der Neuerscheinungen in Deutschland von 2002 bis 2010 © STATISTA	6
1.4	Säurefraß am Buch [Köl12]	7
1.5	Maßnahme zur Haushaltskonsolidierung © STATISTA	8
1.6	Erzeugung von Information [Chu+08]	10
1.7	Prozentanteil in 10 Jahren nicht referenzierter Veröffentlichungen [Les97]	11
1.8	Aufbau eines Emulators	16
1.9	Apple II [Gam12]	18
1.10	Karateka 1986[Gam12]	18
1.11	Bitmuster auf Mikrofilm [Hof12b]	19
1.12	Domesday Book [Tho11]	21
1.13	Domesday System © WIKIPEDIA	22
1.14	Domesday Original	23
1.15	Domesday Emulation	23
2.1	Euro Münze - microSD © WIKIPEDIA	27
2.2	Bibliothek der University of Michigan [Scho5]	27
2.3	Beschädigte CD [Sch12]	30
2.4	Festplattencrash [Dlu12]	30
2.5	Speichermedien - Anteil der Vorfälle von Datenverlust © STATISTA	30
2.6	Man in the middle Attacke © WIKIPEDIA	31
2.7	Verbindung durch HTTPS geschützt	31
2.8	Absatz von Speichermedien auf dem Konsumentenmarkt in Deutsch- land von 2005 bis 2011 © STATISTA	32
2.9	Spezifikation von OOXML in ausgedruckter Form [noo12]	34
2.10	Prognose zum Absatz von eBooks in Deutschland bis 2015 © STA- TISTA	37
2.11	Absatz von E-Readern weltweit von 2009 bis 2011 © STATISTA	38
2.12	Marktanteil von Amazon am Absatz von E-Readern weltweit von 2008 bis 2011 © STATISTA	39

3.1	Dateistruktur einer PDF Datei [Adoo8]	42
3.2	Aufbau eines PDF Dokumentes [Adoo8]	43
3.3	Mögliche Technologien in EPUB 3 (angelehnt an [MW11])	47
3.4	Aufbau eines EPUB Dokumentes	48
3.5	EPUB Ordnerstruktur	49
3.6	Auflistung aller in dem Buch vorhandenen Dateien und Ordner	50
3.7	Editierete EPUB Publikation	54
3.8	Downloadstatistik nach Formaten bei O'Reilly © O'REILLY	57
3.9	Grundlegende DRM Architektur [ORe12a]	60
3.10	Grundlegender Personalisierungsvorgang bei weichem DRM	62
3.11	Wasserzeichen in Musikdatei geladen von iTunes	63
3.12	Suche nach illegalen Werken mit Hilfe von Wasserzeichen [SWD09]	64
3.13	Schutzmaßnahmen von Verlagen gegen Piraterie bei E-Books © STATISTA	66
3.14	Verwendete Metadatenstandards in den Bereichen Bibliothek (orange) und Archiv (violett) [RB10]	67
3.15	MARC Datensatz codiert (Library of Congress Online Catalog)	72
3.16	MARC Datensatz aufbereitet (Library of Congress Online Catalog)	72
3.17	FBFR Entitäten der Gruppe 1 mit Beispiel [Mad+09]	75
3.18	Alle FRBR Gruppen und die Gegenstands (subject) Beziehungen [Mad+09]	76
3.19	Moon+ im Google Playstore	81
3.20	Moon+ Erster Start	81
3.21	Hinzufügen von O'Reilly OPDS	81
3.22	O'Reilly OPDS - Navigation	81
3.23	O'Reilly - Bestseller	82
3.24	Weiterleitung auf O'Reilly Bookstore	82
4.1	Fedora Objekt Modell [DW11]	85
4.2	DSpace Daten Modell [Dig09]	89
4.3	DSpace - Speicherung der Objekte [Kök11]	91
4.4	Repository Systeme [Ber11]	94
5.1	Unterteilung des Layouts mit Hilfe von Divs	97
5.2	Text außerhalb der Div Grenzen	101
5.3	Text von Div Grenze geschnitten	102
5.4	Mit Hilfe von multi-column geladener Text	104
5.5	Verschiebung der einzelnen Spalten	105
5.6	eReader Navigation	105
5.7	eReader mit implementiertem Paging	107
5.8	EPUB Ordnerstruktur	108
5.9	Karateka 1986	108

5.10	eReader mit falschem Pfad zu den Bildern	110
5.11	eReader mit richtigem Pfad zu den Bildern	110
5.12	eReader - Layout mit einer Seite	111
5.13	Fehlerhafte Darstellung der Divs im Internet Explorer	112
5.14	Fehlerhafte Seitenberechnung im Internet Explorer	113

Tabellenverzeichnis

1.1	Zusammenhang zwischen Information und Wissen [Efoo]	3
1.2	Verfügbarkeit von Speichermedien für Langzeitarchivierung © STATISTA	15
1.3	Vergleich von konventioneller und digitaler Bibliothek [MS97]	23
1.4	Aktueller Vergleich von konventioneller und digitaler Bibliothek	25
2.1	Historische Entwicklung von eBooks [Sfe11]	36
2.2	Zusammenhang: Unternehmen, eReader/Tablet, Anzeigeformat	38
3.1	Kontrollmöglichkeiten einzelner Bereiche mittels hartem DRM	58
3.2	Kundenmeinung zur DRM bei E-Books und E-Readern © STATIS- TA	66
3.3	Kernelemente von Dublin Core [dub10]	69
3.4	Grundstruktur von METS [Con12c]	70
3.5	Auszüge von Zahlencodes in MARC) [Con12b]	71
3.6	MODS Top Level Elemente [Con12d]	73
3.7	FRBR Entitäten Gruppen [Mad+09]	74
3.8	OAI Befehlssatz [Lag+02]	77
4.1	Reservierte Datastream Identifier eines Fedora Objekts [DW11]	86
4.2	Beispielunterteilung eines DSpace Objektes [Dig09]	90
5.1	Funktion und Beschreibung der verwendeten Divs	96

Quellcode

3.1	META-INF/container.xml	50
3.2	OPS/epb.opf	51
3.3	OPS/epb.ncx	52
3.4	CSS Quellcode zur EPUB Erweiterung	54
3.5	HTML Quellcode zur EPUB Erweiterung	54
3.6	Beispielhafte Codierung mit Dublin Core in RDF/XML [dub10]	70
3.7	METS Dokument mit Dublin Core Tags [Con12c]	71
3.8	MODS - Codierte Metadaten [Con12d]	73
3.9	Angebotene Metadaten nach einem OAI Request	78
3.10	O'Reilly OPDS Katalog	80
5.1	Ausschnitt einer *.ncx Datei	98
5.2	Auszug einer XHTML Datei	99
5.3	CSS Definition	103
5.4	Einzel gesetzte CSS Pfade	108
5.5	Speziell angepasstes CSS für Internet Explorer	113

Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?

(Thomas Stearns Eliot 1934)

1 Einleitung

Bibliotheken waren schon immer Orte der Information, des Wissenstranfers und der Dokumentation. Mit Beginn des digitalen Zeitalters wurden auch die klassischen Bibliotheken vor neue Aufgaben gestellt. Der Einsatz von Computern in nahezu allen Bereichen der Wissenschaft verhalf diesen zu einer noch nie da gewesenen Wissensgenerierung. Es erfolgt dabei eine vollständige Digitalisierung von Informations- und Kommunikationsprozessen, welche wiederum diesem rasanten Anstieg von Information zu Grunde lagen. Die weltweite Speicherkapazität von Information pro Kopf sind in den zwei Jahrzehnten zwischen 1986 und 2007 zwischen 23% und 28% pro Jahr gewachsen. [HL07]

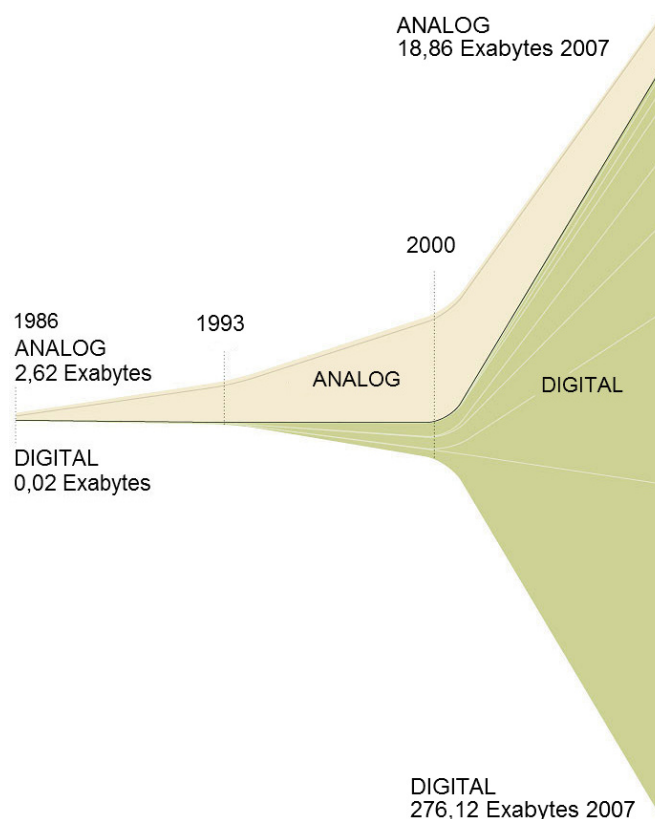


Abbildung 1.1: Speicherkapazität analoger und digitaler Daten [HL07]

Diese Digitalisierung bietet eine Fülle an neuen Möglichkeiten mit dieser Information umzugehen, sie zu benutzen, zu verteilen oder zu publizieren. Es gibt mittlerweile eine große Zahl an digitalen Formaten mit speziellen Vor- und Nachteilen und noch mehr Endgeräte, um diese anzuzeigen.

Die Bibliothek hat somit einerseits die Aufgabe diese neuen Technologien anzubieten und zu verwenden, muss aber andererseits die bis dato gewohnte Qualität ihres Angebotes aufrechterhalten. Durch die Verwendung und Bereitstellung dieser digitalen Formate passiert eine Transformation von klassisch gedruckten Medien hin zu digitalen Dokumenten. Dadurch werden Dokumente seltener als Buch und häufiger als Datei angeboten. Das Bibliotheksgebäude ist bis jetzt ein Ort der Kommunikation und des Lernens bestehend aus Lesesälen und Regalen voller Bücher. Dieser Ort ist mit dem Angebot der Digitalisierung theoretisch nicht mehr notwendig. Ob in Zukunft aus den Regalen mit Büchern eine Cloud wird und der Lesesaal zu einem privaten Platz zu Hause ist eher unwahrscheinlich. Bibliotheken sind immer noch Orte, an denen Wissensaustausch und Kommunikation tatsächlich geschieht, wo sich Menschen kommunizierend begegnen. Sie sind immer noch verantwortlich physischen Zugang zu Information zu bieten, deren Qualität zu gewährleisten und stellen sicher, dass diese über einen längeren Zeitraum sicher aufbewahrt werden.[[Deg12](#)]

Auf den ersten Blick könnte man meinen, dass die Probleme der klassischen Bibliotheken mit dem Übergang zu digitalen Bibliotheken gelöst wurden. Das dem nicht so ist und sich damit auch neue Fragen und Probleme auftun wird in weiterer Folge diese Arbeit zeigen.

1.1 Information - von Daten zur Weisheit

Ein Modell, welches die Generierung von Daten, die daraus resultierende Information und das Wissen bis hin zur persönlichen Erkenntnis darstellt, liefert Nathan Shedroff. Die vier Grundbegriffe dieser Darstellung sind »Data«, »Information«, »Knowledge« und »Wisdom«. In weiterer Folge wurden die Begriffe mit »Daten«, »Information«, »Wissen« und »Weisheit« übersetzt.[[She99](#)]

1.1.1 Daten

In der Informatik versteht man Daten als maschinenlesbare und -bearbeitbare, in der Regel digitale Repräsentation von Information. Der Aufbau solcher Daten erfolgt nach einer bestimmten Syntax, die eindeutige Regeln beinhaltet. Damit

aus Daten Information gewonnen werden kann, muss ein Kontext existieren, um anhand diesem eine Interpretation vornehmen zu können. [Wik12b]

1.1.2 Information

Information ist eine zeitliche Abfolge von Signalen, deren Sinn und Bedeutung der Empfänger nach seinen Möglichkeiten und Fähigkeiten, interpretieren kann. Information wird durch Daten dargestellt. Unterschiedliche Daten bzw. Nachrichten können die selbe Bedeutung haben. Andererseits kann aufgrund von ein und derselben Nachricht je nach Kontext unterschiedliche Information erzeugt werden. [Wik12f]

Zeichenkette	Interpretationsstufe	Bedeutung
28081749	Informationsgehalt	Eine Zahl
	Wissensgehalt Stufe 1	Ein Datum
	Wissensgehalt Stufe 2	Goethes Geburtsdatum

Zeichenkette	Interpretationsstufe	Bedeutung
28081749	Informationsgehalt	Eine Zahl
	Wissensgehalt Stufe 1	Ein Datum
	Wissensgehalt Stufe 2	Breite von Neu Delhi

Tabelle 1.1: Zusammenhang zwischen Information und Wissen [EFoo]

1.1.3 Wissen

Wissen resultiert aus der Interpretation von Daten und Information. Es entsteht durch die Einbeziehung von persönlichen Erfahrungen oder Fremdeindrücken. Wissen soll dadurch zu einer besseren Entscheidungsfindung helfen. [CENo4] Es erfolgt somit eine Unterscheidung dieser beiden Begriffe. Während die Information von außen auf den Betrachter einwirkt, entsteht das Wissen selbst intern beim Betrachter selbst. Im alltäglichen Gebrauch der beiden Begriffe sind diese Grenzen nicht so deutlich gesetzt. Wenn jemand beispielsweise einige Fakten auswendig lernt, wird oft behauptet, dass diese Person viel weiß. Im Grunde ist es aber nur ein Wiedergeben von Information. Wissen resultiert aus der Interpretation und Verknüpfung von Information.

Die Generierung von Wissen ist damit immer einem Prozess unterworfen. In diesem Prozess wird beschafft, analysiert, erfasst, verteilt, gespeichert und auch

vergessen. Es durchläuft damit verschiedene Stadien, kann sich innerhalb von Prozessen in Kreisläufe wieder finden oder dort hängen bleiben. Wichtig dabei ist, dass Wissen erst dann nützlich ist, wenn es von einer Person verarbeitet und verstanden wird. [Bal11]

1.1.4 Weisheit

Weisheit ist die Fähigkeit, das gesammelte Wissen und die Erfahrung zusammenhängend zu verstehen. Die daraus gewonnene Erkenntnis ist das Ergebnis dieser Fähigkeit. Das hier dargestellte Modell zeigt deutlich, dass Daten die Grundlage jeder Erkenntnis bzw. Weisheit sind. Dieser spezielle Prozess von »Daten zur Weisheit« ist ein sehr persönlicher, der beim Betrachter selbst passiert. [Maz09]

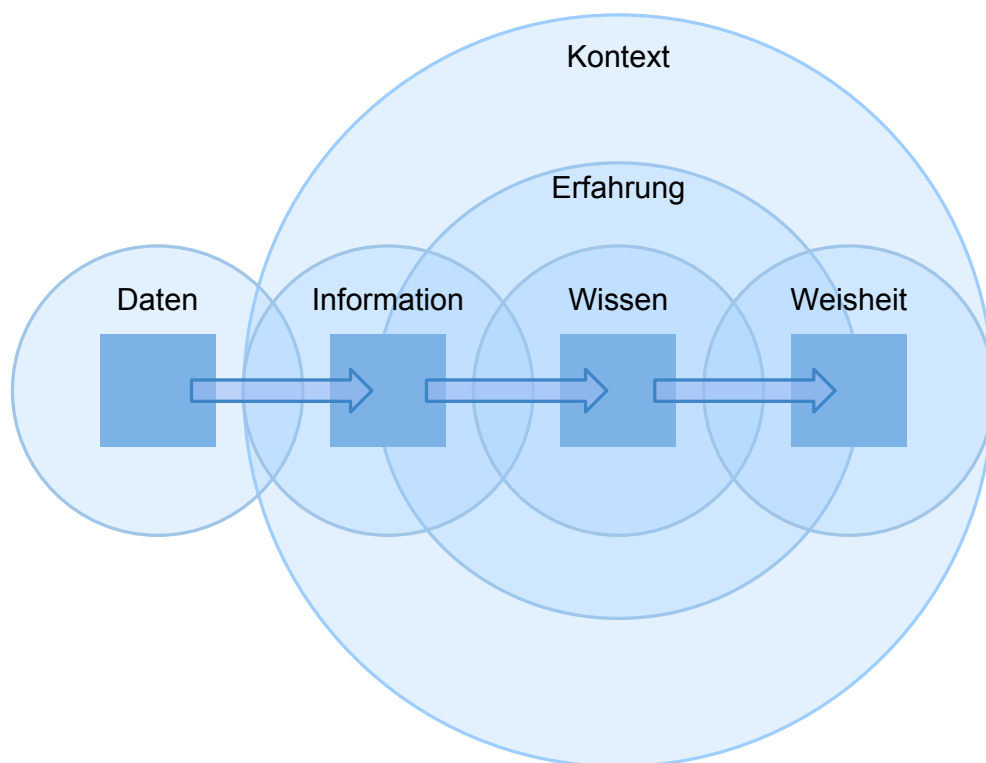


Abbildung 1.2: Prozess: Von Daten zu Wissen nach Nathan Shedroff [She99]

1.2 Konventionelle Bibliothek

Mit konventioneller Bibliothek ist hier die traditionelle Bibliothek gemeint, bestehend aus großen Räumen mit Regalen voll Bücher und Lesesälen. Es ist meist eine öffentliche Einrichtung in der der Nutzer auf eine kategorisierte und geordnete Sammlung von Büchern, Journalen, Zeitschriften oder anderweitige Information physischen Zugriff hat. Des weiteren kann der Nutzer bei seiner Recherche auch auf das Wissen und die Hilfe von Bibliothekaren zurückgreifen. In weiterer Folge dieser Arbeit wird der Begriff »Konventionelle Bibliothek« für Einrichtungen verwendet, in der keine Möglichkeit besteht, digitale Dienste in Anspruch zu nehmen.

1.2.1 Aufgaben von konventionellen Bibliotheken

Das primäre Ziel von Bibliotheken ist die Wissensversorgung. Es wurden dazu fünf aussagekräftige Gesetze definiert. [Ran31]

- Bücher sind zur Verwendung geschaffen
- Jedem Leser sein Buch
- Jedem Buch seinen Leser
- Erspare dem Leser Zeit
- Eine Bibliothek ist ein wachsender Organismus

Bibliotheken sind darauf bedacht, ihren Bestand ständig zu erweitern und zu vergrößern und verfolgen somit das »Alexandrische Prinzip«. Da in den letzten Jahren das Aufkommen an Büchern und Publikationen immer mehr steigt, sind diese Bibliotheken auf der Suche nach neuen Strategien, um dieser Herausforderung Herr zu werden. Dieser Prozess ist ein typischer, langwieriger Beschaffungs- und Erschließungsvorgang. Bibliotheken müssen auch gesetzliche Aufgaben erfüllen, um Publikationen für spätere Generationen zu erhalten. Am Beispiel der Österreichischen Nationalbibliothek ist eine sogenannte Pflichtablieferung erstmals schon ab dem Jahr 1569 dokumentiert. Ab diesem Zeitpunkt wird jedes in Österreich jemals publizierte Dokument in der Nationalbibliothek abgelegt. Das große Ziel dahinter ist eine möglichst lückenlose Dokumentation des publizistischen Schaffens in Österreich und deren Zugänglichkeit für zukünftige Generationen. [Nat12b]

1.2.2 Probleme konventioneller Bibliotheken

1.2.2.1 Lagerprobleme durch mangelnden Platz für Bücher

Durch die ständig steigende Anzahl von Publikationen stehen konventionelle Bibliotheken vor dem Problem, diese Neuerscheinungen auch wo unterzubringen. Es werden laufend neue Lagerräume benötigt, die den speziellen Anforderungen zur Lagerung von Büchern entsprechen müssen. Doch auch wenn die finanziellen Mittel es zulassen neue Lagerräume zu erschließen, so sind diese meist in einiger Entfernung zum Standplatz der Hauptbibliothek. Dieser Umstand verlängert und erschwert wiederum den Ausleihprozess.

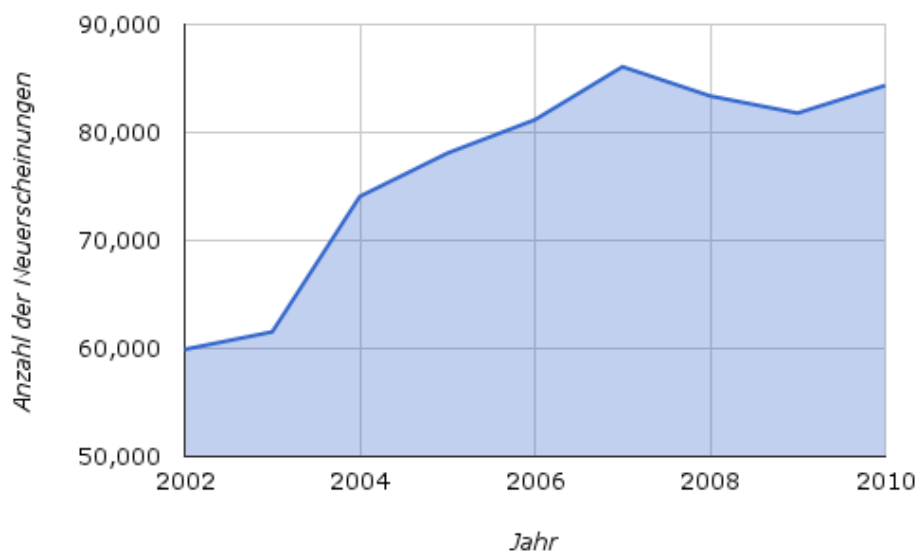


Abbildung 1.3: Anzahl der Neuerscheinungen in Deutschland von 2002 bis 2010 © STATISTA

1.2.2.2 Fehlende Leseräume für Nutzer

Je breiter das Angebot einer Bibliothek ist, desto größer ist auch deren Nutzerstamm. Da die Recherche nicht immer mit dem Ausleihen des Buches verbunden sein muss, sollte dem Nutzer auch ein ruhiger Lesebereich angeboten werden. Gerade im Bereich der Universitäten gibt es jedes Jahr einen Anstieg von Erstsemestrigen und potentiellen Bibliotheksnutzern. Platz gibt es nicht einmal für

einen Bruchteil dieser Nutzer. Erschwerend kommt hinzu, dass viele Studenten die wenigen Räume der Bibliotheken auch als Lernräume zweckentfremden und damit den Platz für Bibliotheksnutzer noch weiter einschränken.

1.2.2.3 Säurefraß und Schimmelbefall an Büchern

Viele Bücher, die Mitte des 19. Jahrhunderts hergestellt wurden, leiden meist an Säurefraß. Das Problem entstand durch die industrielle Herstellung von Papier. Durch die maschinelle Herstellung von Papier entstand Säure, die in weiterer Folge das Grundmaterial von Papier, Zellulose angriff. Es gibt zwar Möglichkeiten, diesen Prozess zu verlangsamen, welche aber sehr kostspielig sind. Berechnungen zur Folge, würde eine Entsäuerung von einem Kilogramm Papier auf 30 Euro kommen. Deutschlandweit sind rund 80 Millionen Bücher von dem Zerfall betroffen. Durch die finanziellen Einschnitte sind die Bibliotheken nun gezwungen, Priorisierungen vorzunehmen, die festlegen, welche Bestände zuerst gerettet werden sollen. Ein Verlust von Büchern durch fehlendes Geld ist somit nicht ausgeschlossen. [Heu03]



Abbildung 1.4: Säurefraß am Buch [Köl12]

1.2.2.4 Distanz von Nutzern

Konventionelle Bibliotheken liegen aus ihrer Geschichte heraus meist zentral in einer Stadt. Der Nutzer muss also immer zu dieser Bibliothek, um deren Angebot zu nutzen. Je weiter ein Nutzer von einer Bibliothek entfernt ist, desto weniger wird sein Wille, diese aufzusuchen. Ein weiterer negativer Faktor sind die beschränkten Öffnungszeiten. Oft sind diese weit weniger flexibel als z.B. der Einzelhandel. Zusätzlich sei aber erwähnt, dass Bibliotheken durchaus die Möglichkeit haben Nutzer an sich zu binden, die finanziellen Mittel vorausgesetzt. Die Stadtbibliothek Augsburg wurde 2009 in einem Neubau feierlich eröffnet mit der Konsequenz, dass die Nutzerzahl binnen eines Jahres von 290.000 auf 526.000 anstieg. [Bib11]

1.2.2.5 Entscheidung zwischen Anschaffung und verfügbarem Budget

Durch die steigenden Kosten der Abonnementspreise von Fachzeitschriften wird die finanzielle Situation von Bibliotheken noch mehr strapaziert.

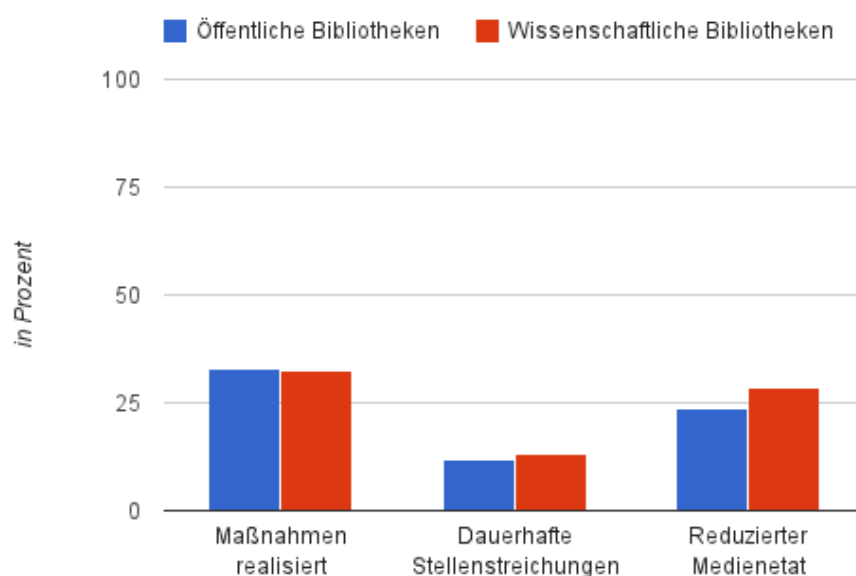


Abbildung 1.5: Maßnahme zur Haushaltskonsolidierung © STATISTA

Im Jahr 2011 war ein Drittel aller Wissenschaftlicher Bibliotheken in Deutschland von Einsparungen betroffen. Viele sind gezwungen das Angebot einzuschränken um kostendeckend arbeiten zu können. Diese Einschränkung an Angebot wirkt sich wiederum auf die Zufriedenheit der Nutzer aus. [Bib11]

1.2.2.6 Dilemma zwischen Breite des Bedarfs und akutem Bedarf

Vor allem Konventionelle Bibliotheken sind aufgrund des notorischen Platzmangels bestrebt, von jedem Exemplar meist nur wenige Kopien zu lagern. Das ist einerseits ein Vorteil, da man das Angebot sehr breit halten kann, ist andererseits aber auch ein Nachteil, da man Interessensspitzen nicht bedienen kann. Wird z.B. in einer Vorlesung einer Universität weiterführende Literatur empfohlen, so ist diese meist danach sehr schnell vergriffen, da in der dazugehörigen Fachbibliothek meist nur drei bis vier Exemplare zur Ausleihe vorhanden sind. [Efoo]

1.2.2.7 Umfangreicher Beschaffungs- und Erschließungsvorgang

Der derzeitige Prozess von der Aquisie bis zur Bereitstellung und Archivierung ist ein sehr langwieriger. Vor allem die Katalogisierung kann bei Publikations-sammlungen sehr zeitraubend ausfallen. Um hier Zeit zu sparen, wird dann meist nur der Name des Journals oder der Tagung angegeben und diese Information in den Bibliothekskatalogen abgelegt. Oft wird auf die Kategorisierung und Beschreibung der einzelnen Beiträge dieser Sammlung oder Tagung verzichtet.[Efoo]

1.2.2.8 Zeitintensive Fernleihe

Da sich viele Bibliotheken nicht immer einen breiten Bestand an Literatur leisten können, sind diese auf weitere Fachbibliotheken angewiesen. Der interessierte Nutzer hat somit die Möglichkeit, ein Buch, das es vor Ort nicht lagernd gibt, von einer anderen Bibliothek zu bestellen. Dieser Vorgang ist leider sehr zeitaufwendig, was dazu führt, dass viele Nutzer in der Zwischenzeit das Interesse an ihrer Bestellung verloren haben. Es wird davon ausgegangen, dass ca. 40% der per Fernleihe bestellten Bücher nicht abgeholt werden. Der Grund dafür ist mit unter die lange Bearbeitungs- und Liferzeit von durchschnittlich 21 Tagen. Die bis dahin angefallenen Kosten bleiben natürlich bei der Bibliothek. [Hir96]

1.3 Digitale Bibliothek

Mit Beginn des digitalen Zeitalters wurde auch der Begriff der digitalen Bibliothek geprägt. In ihr werden digitale Inhalte jeglicher Form aufbereitet und dem Nutzer zur Verfügung gestellt. Die Aufbereitung erfolgt nach festgelegten Regeln, um eine gleichbleibende Qualität gewährleisten zu können. Aufgrund dieser Strukturen können dem Nutzer auch spezielle Dienste angeboten werden. [Can+07]

1.3.1 Entstehung neuer Information

Die in den letzten Jahren rasant angestiegene Informationsflut macht auch vor Bibliotheken nicht halt. Die Bibliotheken heutiger Zeit sind mit steigenden Kosten, neuen Medien, Formaten und dem Wunsch von speziellen Services für den Nutzer konfrontiert. Die steigende Informationsflut basiert zum Teil auf der großen Weiterentwicklung und Akzeptanz des Internets und deren Präsentationsmöglichkeiten. Eine Be- und Verarbeitung dieser ist mit herkömmlichen Mitteln nicht mehr zu schaffen.

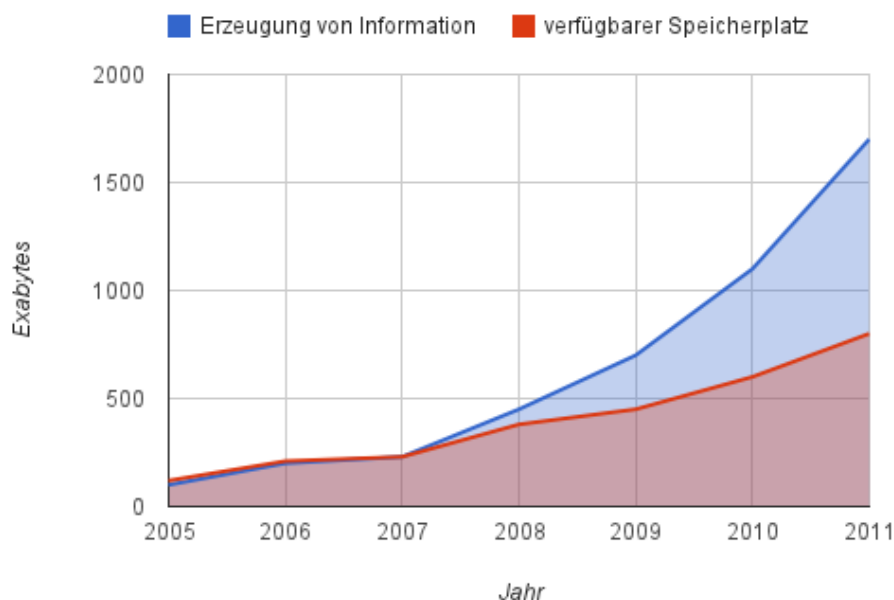


Abbildung 1.6: Erzeugung von Information [Chu+08]

Dass es große Unterschiede in der Wertigkeit dieser Informationen gibt, sieht man an folgender Grafik. Die Wertigkeit der Information ist hier auf die Häufigkeit der Referenzierungen bezogen. An diesem Beispiel wurde festgehalten, welche Publikationen, die im Jahre 1984 publiziert wurden, in den darauf folgenden zehn Jahren nicht referenziert wurden.

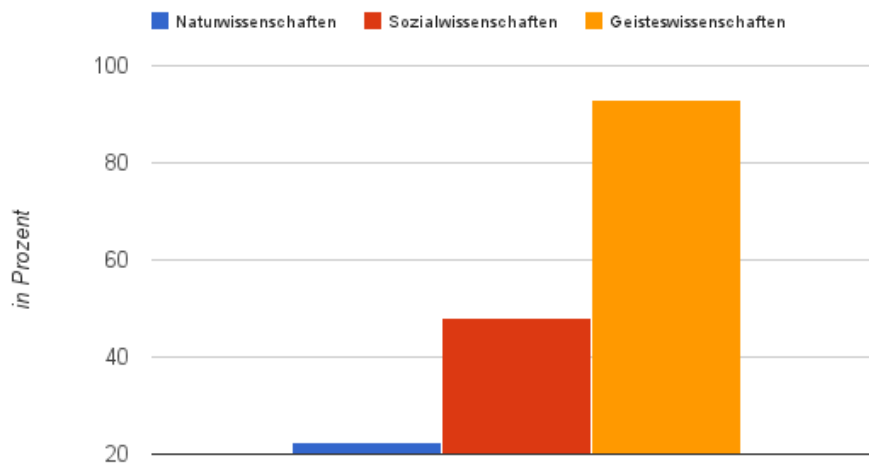


Abbildung 1.7: Prozentanteil in 10 Jahren nicht referenzierter Veröffentlichungen [Les97]

Grund dafür ist der meist große Konkurrenzkampf zwischen den Forschern, um an finanzielle Zuwendung für Projekte zu gelangen. Der Begriff »publish or perish« was so viel bedeutet wie »publiziere oder gehe unter« ist in Bereichen der Wissenschaft eine gängige Redewendung. Oft hängt ein Zuschlag davon ab, wieviel derjenige bereits publiziert hat. Die Folge ist oft ein hoher Druck, um in möglichst vielen unterschiedlichen Journalen seine Beiträge zu veröffentlichen. Dieser Publikationsdruck geht meist auf Kosten der Qualität. [Wik12j]

Um dem entgegenzuwirken, entschied die Deutsche Forschungsgesellschaft bei der Entscheidung von Projektanträgen nur mehr fünf Verweise auf Publikationen entgegenzunehmen. Damit soll der Druck von »publish or perish« genommen werden. Weiters wurde die Empfehlung ausgesprochen, dass Originalität und Qualität stets Vorrang vor Quantität haben muss. [For98]

1.3.2 Aufgaben und Ziele einer digitalen Bibliothek

Es wird versucht, mit digitalen Bibliotheken alle Schwächen und Nachteile von konventionellen Bibliotheken zu lösen. Aus Nutzersicht sollte zum Beispiel der Begriff »vergriffen« der Vergangenheit angehören. Aber auch aus wissenschaftlicher Sicht bietet die digitale Bibliothek aufgrund ihrer Vernetzung jedem einzelnen eine viel spezifischere Möglichkeit an Ressourcen zu gelangen. Es ist dadurch möglich, leichter auf individuelle Bedürfnisse und Tätigkeiten der Nutzer einzugehen. Eine digitale Bibliothek soll auch durch das Bereitstellen von speziellen Diensten und Services überzeugen. So soll es für jede spezifische Gruppe von Nutzern, egal ob privat oder wissenschaftlicher Natur ein Leichtes sein, effizient an die von ihnen gewünschte Information zu kommen und diese auch in mehreren Formaten betrachten zu können. Diese Aufgaben sind am Beispiel der Österreichischen Nationalbibliothek gesetzlich geregelt. Laut § 3, Abs.1 der Bibliotheksordnung der Österreichischen Nationalbibliothek (Verordnung vom 11. Jänner 2002, BGBl 12/2002) zählt die Sammlung und Archivierung aller in Österreich publizierten Medien einschließlich der elektronischer Medien (§§ 43 und 43a Mediengesetz) zu ihren wesentlichen Aufgaben. Durch die steigende Digitalisierung und die dadurch geänderte Medienlandschaft wird aber immer mehr online publiziert. Diese Medien waren bis 2009 von der Pflichtablieferung nicht betroffen. Dieser Umstand führte Anfang 2009 zu einer Novelle des Mediengesetzes. Diese Novelle war dahingehend notwendig, um etwaigen Dokumentationslücken vorzubeugen. Dadurch wurden auch hier alle Aspekte zur Ablieferung elektronischer Medien geregelt. Nach (§ 43b Mediengesetz) darf die Österreichische Nationalbibliothek maximal vier mal im Jahr den Inhalt periodischer elektronischer Medien automatisch sammeln. Primär liegt der Schwerpunkt hier auf Text und Bildmaterial. Des weiteren muss ein Bezug zu Österreich aufgewiesen werden und die Information muss unter einer »at« Domäne abrufbar sein. [Nat12a]

1.3.3 Produkte einer digitalen Bibliothek

Das primäre Produkt einer digitalen Bibliothek sind ihre Dokumente in digitaler Form. Für die Aufschlüsselung und Beschreibung der gebräuchlichsten und weit verbreitetsten Dokumente siehe Abschnitt 3.1 auf Seite 41. Ein weiterer wichtiger Bestandteil sind Metadaten, die zu jedem Dokument vorhanden sein sollen. Mit dieser Information ist eine schnelle automatisierte Kategorisierung der Inhalte möglich. Mehr dazu siehe Abschnitt 3.4 auf Seite 67. Des weiteren werden durch die hohe Vernetzung der Kommunikationswege spezielle Dienste angeboten, die dem Nutzer eine noch umfangreichere Recherche erlauben.

1.3.4 Probleme digitaler Bibliotheken

Die Verwendung digitaler Medien bringt auch deren Nachteile mit sich. Mehr dazu siehe Abschnitt 2.1.2 auf Seite 28. Das weitaus kostenintensivste Problem digitaler Bibliotheken ist die Langzeitarchivierung. Da es eine Fülle an Formaten siehe Abschnitt 3.1 auf Seite 41 gibt, muss auch eine Verwendung dieser in Zukunft gewährleistet werden. In den letzten Jahren hat es nicht nur im Hardwarebereich einen enormen Technologiewandel gegeben. Neue Geräte und Funktionen setzen auch neue Dokumentenformate voraus, die sukzessive alte Standards ablösen. Es ist somit eine große Herausforderung, dieses in alten Formaten publizierte Wissen weiter zugänglich zu halten. Zusätzlich zu den alten Formaten muss auch die dazugehörige Hardware verfügbar sein. Alte Software muss oder kann nicht immer auf neuer Hardware laufen.

1.3.5 Langzeitarchivierung

Es gibt verschiedene Ansätze die Zugänglichkeit digitaler Medien zu gewährleisten. Der gemeinsame Aspekt dieser Ansätze ist eine langfristige Finanzierung. Wie hoch diese in Zukunft ausfallen werden, kann trotz aller Planung nur teilweise abgeschätzt werden. Da diese Aufgabe großteils öffentlichen Institutionen vorbehalten bleibt und diese generell mit finanziellen Einschnitten zu kämpfen haben, ist ein Zusammenschluss einzelner digitaler Bibliotheken sicherlich von Vorteil. [JS12]

1.3.5.1 Bitstream Preservation

Bei dieser Technik geht es darum, den Bitstream, also den Datenstrom zu erhalten. Diese Strategie ist in zwei Schritten aufgebaut. Einerseits muss der Bitstream auf ein Medium gespeichert werden und gleichzeitig muss die geeignete Hardware und Software vorhanden sein, um dieses Medium lesen zu können. Grundvoraussetzung dafür ist der unbeschädigte Bitstream. Nun gibt es mehrere Ansätze von Speicherstrategien, die einen fehlerfreien Bitstream gewährleisten sollen.

Redundante Datenhaltung: Die Daten sollten mehrfach kopiert werden. Um auch bei der Lagerung der Medien auf Nummer sicher zu gehen, sollten die Kopien an jeweils unterschiedlichen Orten gelagert werden. Durch diese Vorgangsweise kann ein Verlust durch äußere Einflüsse wie z.B. Feuer oder Wasserschaden minimiert werden. Um die Sicherheit noch weiter zu erhöhen, kann der gesamte Prozess der Speicherung organisatorisch und

technisch an zwei voneinander unabhängige Organisationen vergeben werden. Damit wird das Risiko von menschlichem und technischem Versagen weitaus geringer.

Divergente Speichermedien: Die gesicherten Daten sollten auf mindestens zwei unterschiedlichen Speichermedien abgelegt werden. Die Datenträger sollten eine breite Verwendung finden und sich am Markt etabliert haben.

Verwendung von Standards: Bei der Speicherung sollte darauf geachtet werden, dass weit verbreitete, gut dokumentierte und anerkannte Formate und Standards verwendet werden. Wenn möglich, sollte auch eine Fehlerkorrektur unterstützt werden.

Regelmäßiger Tausch der Medien: Die verwendeten Speichermedien sollten in regelmäßigen Abständen durch neue ersetzt werden. Dadurch sollte ein Verlust von Daten durch fortschreitende Alterung vermieden werden. Unmittelbar nach der Speicherung sollte auch die Lesbarkeit der Medien überprüft und in Form einer Fehlerstatistik kontrolliert werden. Des Weiteren sollte auf eine fachgerechte Lagerung geachtet werden.[Ull12]

1.3.5.2 Migration

Um eine Migration durchführen zu können, ist ein Medium und ein unbeschädigter Datenstrom Voraussetzung. Sie muss durchgeführt werden, wenn Formate oder Datenträger Gefahr laufen nicht mehr unterstützt zu werden. Eine Migration von digitalen Objekten kann zwei verschiedene Prozesse betreffen, einerseits die Datenträgermigration und andererseits die Datenmigration. Bei der Datenträgermigration passiert ein Tausch des Trägermediums. Beispielsweise kann hier Information, die auf Festplatte vorliegt, auf DVD/Blu-ray Disc oder Band gespeichert werden. Diese Art von Migration ist noch relativ schnell und einfach zu erledigen. Etwas komplizierter ist die Datenmigration. Bei der Datenmigration wird das Format selbst geändert. Voraussetzung dafür ist eine sehr genaue Beschreibung und Kenntnis des Quellformates. Jede noch so kleine Eigenschaft des Quellformates muss vom Nachfolgeformat interpretier- und darstellbar sein. Bei vielen Formaten ist die Information selbst nur ein kleiner Teil, der Rest besteht aus Steuer- und Formatierungsinformation. [Fun12b]

Am Beispiel von Microsoft Word kann das Problem der Migration in einem kleinen Rahmen nachvollzogen werden. Wird hier versucht ein älteres Dokument (z.B. Word 95) in einem aktuellen Office zu öffnen, geschieht dies oft nur über einen internen Konvertierungs- bzw. Migrationsschritt. Zusätzlich verwendet Microsoft Office ab der Version 2007 ein vollkommen neues auf XML

(Extensible Markup Language) basierendes Standard-Dateiformat mit dem Namen OpenXML. Die Spezifikation von OpenXML umfasst über 7000 Seiten und würde daher als Zielformat einer Migration wohl kaum zum Einsatz kommen. [ISOo8]

Bezüglich der Langzeitarchivierung von digitalen Objekten wurde Anfang 2009 eine Umfrage an Experten für Internet, Kommunikation und Medien durchgeführt. Es ging dabei um die Entwicklung von speziell für die Langzeitarchivierung geeigneten Datenträgern. Die Fragestellung lautete wie folgt:

»Was glauben Sie, wann neue Speichermedien für eine Langzeitarchivierung von Dokumenten auf dem Markt verfügbar sein werden?«

	2010 bis 2014	2015 bis 2019	2020 bis 2024	2025 bis 2030	später als 2030	Wahr- scheinlich nie
Deutschland-Experten	24	42	15	9	8	3
EU-Experten	25	25	25	25	0	0
DNAdigital-Experten	27	30	15	12	3	12

Tabelle 1.2: Verfügbarkeit von Speichermedien für Langzeitarchivierung © STATISTA

Auch unter Experten ist die Meinung dazu sehr unterschiedlich. Besonders im Bereich Computer und Informationstechnologie sind Prognosen, die über 10 Jahre in die Zukunft gehen, schwierig. Ein Blick in die Vergangenheit zeigt, dass eine breite Verwendung von Datenträgern meist ca. 20 Jahre anhält, bevor diese von neuen Technologien abgelöst werden. Die Haltbarkeit dieser Medien ist zwar meist höher als deren Verwendungszeitraum, trotzdem fehlt auf lange Sicht die Hardware, um diese lesen zu können.

Vorteile

- Migration kann oft automatisiert durchgeführt werden
- Migration ist nach dem Analysevorgang sehr gut technisch umsetzbar
- Es bleiben Originalobjekte für eine weitere Verwendung zurück (Backup)

Nachteile

- Jeder Datenträger und jedes Objekt muss einzeln migriert werden. (zeit-aufwendig)

- Die Wahrscheinlichkeit von partiellem Datenverlust oder Datenveränderung ist relativ hoch
- Quell- und Zielobjekt jeder Migration sollte gespeichert werden (hoher Speicherplatzbedarf)
- Die Vorbereitung für eine Migration ist sehr zeitintensiv (Analyse der Formate)
- Migration kann nicht immer eingesetzt werden (proprietäre, nicht gut dokumentierte Formate)

1.3.5.3 Emulation

Bei einer Emulation bleiben die Ursprungsformate und Datenträger unverändert. Es wird dabei versucht die Umgebung selbst nachzustellen. Emulation kann zwei verschiedene Prozesse betreffen. Einerseits eine Emulation von Software und andererseits eine Emulation von Hardware. Bei der Emulation von Hardware wird diese mittels Software nachgebildet. Ein Beispiel dazu wäre DOSBox. DOSBox emuliert die Hardware früherer IBM PCs. [Vee+12]

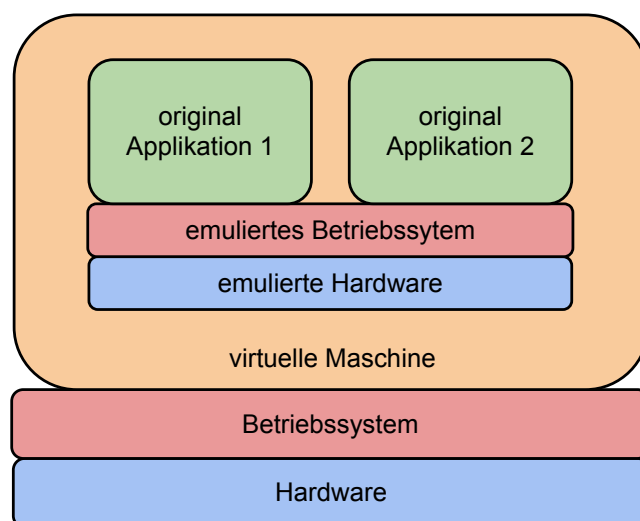


Abbildung 1.8: Aufbau eines Emulators

Programme oder Spiele aus der DOS Zeit wären auf heutigen Rechnern nicht mehr lauffähig. Durch diesen Emulator können aber weiterhin Programme bearbeitet und deren Inhalt zugänglich gemacht werden. Die Information selbst

bleibt dabei in den ursprünglichen Dokumentenformaten gespeichert. Um alte Dokumente anzeigen zu können, muss nicht immer zwingend die gesamte Hardwareumgebung emuliert werden. Es genügt hier oft das Anwendungsprogramm selbst zu emulieren. Sollte das Originalprogramm nicht mehr verfügbar sein oder auf neuer Hardware laufen, kann ein neues Programm geschrieben werden, welches die Funktionalität des alten nachbildet. Ein Beispiel dafür wäre das Programm LibreOffice, welches auch ältere Microsoft Office Dateien problemlos öffnen, bearbeiten und diese auch wieder im original MS Office Dateiformat speichern kann. [Fou12c]

Hier würde auch der Vorteil einheitlicher Standards und Dokumentenformate zum Tragen kommen. Diese würden von vielen Programmen verwendet und damit die Wahrscheinlichkeit minimieren nicht mehr gelesen werden zu können. [Fun12a]

Vorteile

- Die Originalobjekte bleiben erhalten und werden nicht verändert
- Keine Datenmigration notwendig
- Da keine Kopien erstellt werden, ist auch der Speicherplatz geringer

Nachteile

- Emulatoren sind schwer zu implementieren. Je höher die Funktionalität desto größer der Implementierungsaufwand
- Für jede neue Hardware (Hostsystem) muss ein neuer Emulator geschrieben werden
- Die Dokumentation der zu emulierenden Hard- oder Software ist nicht immer hinreichend bekannt oder offen gelegt

1.3.5.4 Computermuseum

Wenn noch keine Strategien zur Migration oder Emulation vorhanden sind, muss auf die Erhaltung der bestehenden Hardware und Software zurückgegriffen werden. Dieser Weg wird zur Langzeitarchivierung nicht empfohlen, da doch einiges an Aufwand und auch Risiko offen bleibt. Bei der Erhaltung von Hardware wird hier zwischen zwei Arten unterschieden. Einerseits die Erhaltung um Digitale Objekte anzuzeigen und andererseits die Erhaltung im Kontext

eines Museums. Zweiteres ist noch aufwendiger, da auch mögliche Reparaturen möglichst historisch korrekt durchgeführt werden müssen. Hinzu kommt noch die begrenzte Lebensdauer der verbauten Bauteile. Vor allem die oft verwendeten Elkos (Elektrolykondensatoren) bilden hier einen Schwachpunkt. Diese beinhalten ein flüssiges Elektrolyt was im Laufe der Zeit durch das Bauteil diffundiert. Dadurch verringert sich die Funktion des Elkos was in weiterer Folge bis zum Ausfall des ganzen Gerätes führen kann. Die Lebensdauer ist somit von Betriebsdauer, Temperatur und der daraus folgenden Verdunstungsrate abhängig. [Hut12]

Ein Beispiel wo diese Art Archivierung angewendet wird, ist das Computerspielmuseum in Berlin. Dieses 1997 eröffnete Museum bietet mit über 300 Exponaten eine Experimentierlandschaft zur Kulturgeschichte des Spiels. Der Bestand vom Januar 2012 beinhaltet ca. 22000 Computerspiele und Anwendungen über 300 Konsolen und Computersysteme und weit über 10000 Zeitschriften. Dazu kommen noch Automatenysteme, Literatur, Medienkunstobjekte und Videos. Das Museum ist weiters Mitglied in der in Deutschland vom Bundesministerium für Bildung und Forschung finanzierten Projekt »NESTOR«. Dieses Projekt ist ein Kompetenznetzwerk zur digitalen Langzeitarchivierung. [Gam12]



Abbildung 1.9: Apple II [Gam12]



Abbildung 1.10: Karateka 1986 [Gam12]

Vorteile

- Erhaltung des ideellen Wertes vor allem über die Hardware
- Erhaltung kann vor allem während der Entwicklung von Emulatoren von Vorteil sein. (um die Neuimplementierungen vergleichen zu können)

Nachteile

- Großer Platzbedarf bei der Lagerung

- Hoher Aufwand für Instandhaltung und Reparaturen
- Begrenzte Lebenserwartung von Medien und Hardware (Elektroteile, Gummiteile, Plastik)
- Hoher Aufwand um die Geräte fachgerecht zu lagern (Temperatur, Staub)

1.3.5.5 Mikroverfilmung

Im Vergleich zu Papier haben digitale Datenträger eine verschwindend geringe Haltbarkeit. Während Papier Information über hunderte von Jahren zugreifbar hält sind das bei digitalen Medien nur einige wenige Jahrzehnte. Gute Erfahrungen in diesem Bereich bietet der Mikrofilm. Dieser soll laut Untersuchungen bis zu 500 Jahre halten. Der Mikrofilm ist eine sehr einfache Möglichkeit Information zu archivieren. Ein Problem ist dabei die Speicherung von Metadaten. Diese müsste wiederum auf eine andere Weise erfolgen. Es gibt Überlegungen den Mikrofilm als Zwischenlösung zur Bitstream Preservation zu verwenden. Es würde dann durch Abfolge von Nullen und Einsen ein sogenanntes Bitmuster gespeichert werden. Ziel dabei ist es die Entwicklung eines geeigneten digitalen Datenträgers abzuwarten. Die Daten auf Mikrofilm könnten dann redigitalisiert und auf dem neuen Medium gespeichert werden. [Kei12]



Abbildung 1.11: Bitmuster auf Mikrofilm [Hof12b]

Vorteile

- Daten sind sehr lange haltbar und aufrufbar
- Datenverlust ist bei richtiger Lagerung gering (kühle Umgebung)
- Aufwand für Lagerung ist gering
- Daten sind nicht veränderbar

Nachteile

- Kategorisierung und Zuweisung von Metadaten schwierig und aufwendig
- Die Verwendung ist an spezielle Lesegeräte gebunden

1.3.6 Domesday Projekt

Grundlage für dieses Projekt war das originale Domesday Book, welches im Jahre 1086 auf Veranlassung Wilhelms des Eroberers geschaffen wurde. Dieses Buch regelte die durch Eroberung und Enteignung entstandenen Lehnverhältnisse. Weiters gab das Buch einen guten Überblick vom Leben der damaligen Zeit. Jede Bescheinigung über Landbesitz hatte dieses Buch als Grundlage. Dieses Buch ist auch noch über 1000 Jahre nach dessen Erstellung Nachschlagewerk und Quelle von Historikern. [Wik12d]



Abbildung 1.12: Domesday Book [Tho11]

Wie wichtig Überlegungen zur richtigen Speicherung digitaler Güter sind, zeigt ein Beispiel aus der Vergangenheit. Im Jahr 1984 startete die BBC (British Broadcasting Corporation) ein Projekt mit dem Ziel eine Bestandsaufnahme vom alltäglichen Leben in England zu erstellen. Dies sollte Texte, kurze Filmsequenzen, Bilder und Tondokumente beinhalten. Es wurde dazu ein Aufruf an die Bevölkerung gestartet, diese Daten zu sammeln und an die BBC zu übermitteln. An diesem Projekt nahmen über 1 Million Menschen teil. [Wik12a]

Die damals verwendete Technologie war die beste ihrer Zeit. Die gesammelten Daten wurden auf zwei Laserdisks gespeichert, die pro Seite eine Speicherkapazität von jeweils 300MB hatten. Als Format wurde das LaserVision Read Only Memory (LV-ROM) verwendet. Gelesen wurden die Laserdisks mit einem dafür speziell entwickelten Player VP415 »Domesday Player« von Philips. Die Software dazu wurde in BCPL (Basic Combined Programming Language) geschrieben. Ziel war es, die gesammelten Daten Schulen und Bibliotheken zur Verfügung zu stellen. Ein weiteres Ziel war es, die gesammelten Daten auch späteren Generationen zur Verfügung zu stellen ähnlich dem Domesday Book. Da neueste

Technik verwendet wurde, war der hohe Anschaffungspreis für viele Einrichtungen aber nicht leistbar. [BBC12]

Der veranschlagte Preis eines Domesday Systems lag bei Projektbeginn bei £1100. Bei Projektende und beim Marktstart lag dieser jedoch bei über £4000. Damit war das Domesday Projekt ein kommerzieller Misserfolg und fand auch nicht die angestrebte Verbreitung. [Melo3]



Abbildung 1.13: Domesday System © WIKIPEDIA

Nur ca. 15 Jahre später war die Anzahl der Player, die diese Laserdisks lesen konnten so gering, dass die Gefahr bestand die Daten zu verlieren. Um dem entgegenzuwirken, wurde das Projekt CAMiLEON (Creative Archiving at Michigan & Leeds) beauftragt mit dem Ziel die gesammelten Daten zu extrahieren und wieder lesbar zu machen. Als einer der ersten Schritte wurde versucht den Bitstream von den Laserdisks auf aktuelle Hardware zu kopieren. Mittels eines Linux Rechners, der mit dem Laserdiskplayer verbunden wurde, gelang es die Datenbank und die Texte auszulesen. Bilder und Videosequenzen wurden mit einem Videograbber mit maximal möglicher Auflösung ausgelesen und unkomprimiert gespeichert. Pro Seite Laserdisk wurden mit dieser Technik ca. 70GB an Daten gesichert. Der weit größere Aufwand bestand darin, die veraltete Hardware zu emulieren. Zu emulieren war der BBC Master Computer, der Laserdiskplayer und die SCSI (Small Computer System Interface) Kommunikation. Bei

der Entwicklung des Emulators wurde darauf Wert gelegt diesen so unabhängig wie möglich zu gestalten, um ihn auch auf zukünftigen Computersystemen starten zu können. Nur durch sehr großen Aufwand war es möglich die Daten des Domesday Projektes wieder verfügbar zu machen.[Mel03]



Abbildung 1.14: Domesday Original

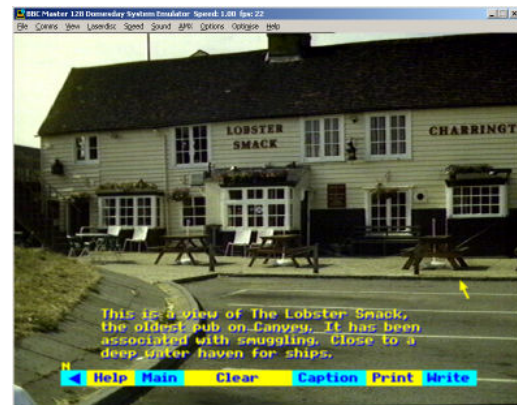


Abbildung 1.15: Domesday Emulation

1.4 Vergleich von konventioneller und digitaler Bibliothek

Konventionelle Bibliothek	Digitale Bibliothek
Hohe Kosten	Niedrige Kosten
Ortsgebunden	Nicht ortsgebunden
Feste Öffnungszeiten	Flexible Öffnungszeiten
Großer Platzbedarf	Geringer Platzbedarf
Dokumente u.U. nicht verfügbar	Dokumente immer verfügbar
Basiert auf örtlichem Bestand	Kann verteilte Bestände kombinieren
Instandhaltungs- und Ersatzprobleme	Wenige Instandhaltungsprobleme
Manuelles Indexieren und Suchen	Automatisches Indexieren & Suchen
Originale	Reproduktionen
Angenehm zu lesen	Weniger angenehm zu lesen
Im Kontext stehend	Aus dem Kontext gerissen

Tabelle 1.3: Vergleich von konventioneller und digitaler Bibliothek [MS97]

Eine detaillierte Gegenüberstellung von konventionellen und digitalen Bibliotheken wurde bereits von Mitchell und Strimpel in einem Artikel im Jahr 1997

erstellt. [MS97] In diesem Vergleich liegt der Schwerpunkt eher auf die gleichzeitige Nutzung von Inhalten. Zum Zeitpunkt der Erstellung sprechen gut zwei Drittel der Punkte für eine digitale Bibliothek. Nach neuesten Erfahrungen und Erkenntnissen im Bereich der digitalen Bibliotheken sind einige dieser Punkte diskussionswürdig.

Aufgrund von technologischen Fortschritten und neu hinzu gekommenen Problemstellungen sind folgende Vor- und Nachteile nicht mehr aktuell.

1.4.1 Hohe Kosten - Niedrige Kosten

Die Instandhaltungskosten von digitalen Bibliotheken werden in Zukunft sehr wohl steigen. Vor allem der Bereich der Langzeitarchivierung LABEL ist ein Punkt, welcher sehr kostspielig sein kann. Des weiteren sind auch die Abonnements von Fachzeitschriften in elektronischer Form jedes Jahr von Preissteigerungen betroffen. Diese automatische Steigerung beträgt pro Jahr ca. 5%. Gerade im universitären Bereich, wo Bibliotheken oft von Kürzungen betroffen sind, ist dieser Umstand problematisch. Vor allem bei Abonnements und Journalen geben Universitäten bis zu 50% ihres Budgets aus. Auf technischen Universitäten kann dieser Wert bis zu 85% betragen. Diese Preissteigerung ging soweit, dass Anfang des Jahres 2012 der renommierte Fachverlag Elsevier von Forschern boykottiert wurde. Kritisiert wurden vor allem die überhöhten Abonnementgebühren und die Praxis der Paketverkäufe. [der12]

1.4.2 Instandhaltungs- und Ersatzprobleme - Wenige Instandhaltungsprobleme

Die hinter einer digitalen Bibliothek stehende Infrastruktur ist sehr wartungsintensiv. Da es Abhängigkeiten von Hardware und Software gibt, muss Sorge getragen werden, dass deren Funktionalität immer gewährleistet ist. Allein der Ausfall einer wichtigen Netzwerkkomponente kann die Dienste einer digitalen Bibliothek vollkommen still legen. Zusätzlich ist weiteres Fachwissen für den Betrieb der Infrastruktur notwendig. Dieses muss, wenn nicht selbst angeeignet, eingekauft werden, was wiederum mit Kosten verbunden ist.

1.4.3 Angenehm zu lesen - Weniger angenehm zu lesen

Vor allem die Entwicklung des elektronischen Papiers kurz E-Paper hat wesentlich dazu beigetragen, dass damit hergestellte Displays sehr gut zum Lesen geeignet sind. Sie bieten sogar Vorteile gegenüber herkömmlichem Papier, was die Leseigenschaften betreffen. So sind diese auch unter sehr heller Umgebung noch gut zum Lesen geeignet. [Wik12e] Einer der ersten E-Reader mit dieser Technologie war der von Amazon entwickelte und sehr weit verbreitete Kindle.

Konventionelle Bibliothek	Digitale Bibliothek
Ortsgebunden	Nicht ortsgebunden
Feste Öffnungszeiten	Flexible Öffnungszeiten
Großer Platzbedarf	Geringer Platzbedarf
Dokumente u.U. nicht verfügbar	Dokumente immer verfügbar
Basiert auf örtlichem Bestand	Kann verteilte Bestände kombinieren
Manuelles Indexieren und Suchen	Automatisches Indexieren & Suchen
Originale	Reproduktionen
nicht von Technologien anhängig	sehr stark an Technologien gebunden
Erfahrung mit Langzeitarchivierung	keine Erfahrung mit Langzeitarchivierung
wenig fachfremdes Wissen notwendig	zusätzliches Wissen notwendig (Instandhaltung, Bedienung,...)
keine Probleme mit Lizenzierung	Probleme mit Lizenzierung

Tabelle 1.4: Aktueller Vergleich von konventioneller und digitaler Bibliothek

2 Grundlagen

The change from atoms to bits is irrevocable and unstoppable. Why now? Because the change is also exponential — small differences of yesterday can have suddenly shocking consequences tomorrow.

(Nicholas Negroponte 1995)

2.1 Digitale Dokumente

Die Inhalte digitaler Bibliotheken werden über digitale Dokumente zur Verfügung gestellt. Digitale Dokumente können in verschiedenen Formen auftreten. Diese Formen können sein: Text, Ton, Film, Bild, Pläne, Animationen oder Kombinationen aus diesen. Digitale Dokumente sind digital kodiert, auf einem Datenträger gespeichert und benötigen spezielle Software und Hardware um verwendbar zu sein.

2.1.1 Vorteile digitaler Dokumente

Digitale Dokumente haben gegenüber analogen Dokumenten wie Büchern, Zeitungen, Skizzen, Fotos oder Tonträgern wesentliche Vorteile. Diese Vorteile spiegeln sich auch in der Gegenüberstellung von digitaler Bibliothek und konventioneller Bibliothek wieder. Sie sind elektronisch verarbeitbar, leicht zu transportieren und werden auf digitalen Datenträgern gespeichert. Die Speichertechnologien sind hierbei magnetisch, optisch oder elektronisch. Folgende Vorteile werden nun gesondert beschrieben.[Efoo]

2.1.1.1 Speicherkapazität

Digitale Dokumente lassen sich in großer Anzahl auf die derzeit verfügbaren Datenträger speichern. Diese Datenträger haben eine so große Kapazität, dass sich z.B. auf einer microSD Karte eine Unmenge an Büchern speichern lassen. Die maximale Kapazität einer microSD Karte liegt derzeit bei 2TB. Wenn man davon ausgeht, dass ein Buch in digitalisierter Form mit ca. 150 Seiten und Illustrationen 1MB an Speicherplatz benötigt, könnten auf eine dieser 2TB großen Speicherkarten ca. 2.000.000 Bücher gespeichert werden. Dieses Rechenbeispiel soll den ungeheuren Platzvorteil digitaler Dokumente zeigen.[Efoo]



Abbildung 2.1: Euro Münze - microSD © WIKI-PEDIA
Abbildung 2.2: Bibliothek der University of Michigan [Scho5]

2.1.1.2 Ausleihe und Übertragung

Wurde ein Dokument gefunden, so ist es bei der Bandbreite heutiger Netzwerke innerhalb von Sekunden lokal verfügbar. Diese Verfügbarkeit ist unabhängig von Öffnungszeiten und Standorten. Egal, wo auf der Welt sich dieses Dokument befindet, ist es für den Nutzer innerhalb kürzester Zeit zugreifbar. Dieser Umstand ist ein riesiger Vorteil gegenüber der z.B. langwierigen Fernleihe.[Efoo]

2.1.1.3 Verfügbarkeit von Exemplaren

Im Vergleich zu herkömmlichen Büchern kann die digitale Version nie vergriffen sein. Vor allem im universitären Sektor ist diese Eigenschaft sehr von Vorteil. Denn gerade hier müssen sich oft hunderte Studenten einige wenige Exemplare einer Universität teilen. Durch die Digitalisierung des Angebots ist ein paralleles Nutzen dieser Ressourcen möglich.[Efoo]

2.1.1.4 Selektive Verwendung

Gerade nach Konferenzen werden die vorgestellten Themen in Sammlungen publiziert. Ist der Nutzer aber nur an einem Artikel interessiert, muss er sich bis dato die gesamte Sammlung ausleihen. Durch die Digitalisierung ist dieser Artikel erstens früher verfügbar und zweitens nicht an eine Sammlung gebunden. Dieser Vorteil lässt auch flexible Abonnementmodelle zu, die vor allem den Nutzern entgegenkommen würden, die nur für das bezahlen wollen, was sie auch interessiert. [Efoo]

2.1.1.5 Weiterverwendung und Reproduzierbarkeit

Ein digitales Dokument kann, sofern die Sicherheitseinstellungen es zulassen, beliebig bearbeitet werden. Voraussetzung dafür ist eine Software, welche eine Bearbeitung des Formates unterstützt. Weiters können Teile z.B. Bilder aus einem Dokument entnommen und einem anderen hinzugefügt werden. Des Weiteren ist es sehr einfach digitale Dokumente zu reproduzieren. Im Zuge der weiteren Bearbeitung können diese auch gemeinsam von mehreren Nutzern gleichzeitig modifiziert werden. [EFoo]

2.1.1.6 Erweiterte Erschließbarkeit

Ein digitales Dokument kann besser aufbereitet werden als analoge Dokumente. Dies kann durch Metainformation oder durch die Analyse und Bewertung des gesamten Inhalts geschehen. Es ist dadurch möglich mit Hilfe von Zusatzprogrammen ein Dokument in Echtzeit zu konvertieren z.B. eine Audioausgabe des Textes oder Aufbereitung in Blindenschrift. Mit der richtigen Software ist es möglich auch Anmerkungen zum Inhalt zu verfassen oder diese hervorzuheben. [EFoo]

2.1.1.7 Kombination unterschiedlicher Darstellungsmöglichkeiten

Mit Hilfe digitaler Dokumente ist es möglich mehrere Arten von Darstellungsmöglichkeiten miteinander zu kombinieren. Es können dadurch auch interaktive Dokumente erstellt werden. Es kann somit im Text ein interner Verweis auf eine Audiodatei erfolgen, die bei Bedarf abgespielt werden kann. Des Weiteren können im Dokument auch z.B. umfangreiche Statistiken in Form von Animationen dargestellt werden, um das Verständnis zu erleichtern. [EFoo]

2.1.2 Nachteile digitaler Dokumente

Bei all den genannten Vorteilen gibt es auch ein paar Nachteile, die nicht außer acht gelassen werden können. Es folgen nun einige Punkte, deren Gewichtung aber sehr stark vom Blickwinkel des Nutzers abhängen können.

2.1.2.1 Abhängigkeit von Software und Hardware

Der vielleicht größte Nachteil von digitalen Dokumenten ist deren Abhängigkeit von spezieller Software und Hardware. Um digitale Dokumente nutzen zu können, muss eine Softwareumgebung existieren, die auf einer dafür ausgelegten Hardware läuft. Aufgrund der unterschiedlichen digitalen Formate kann auch die Betrachtungssoftware sehr umfangreich werden. Um dann die oben genannten Vorteile wie z.B. »Anmerkungen verfassen« nutzen zu können, muss auch der Umgang mit diesen Programmen vorausgesetzt werden. Es sind zumindest Grundkenntnisse notwendig. [Efoo]

2.1.2.2 Leichte Veränderbarkeit

Werden digitale Dokumente nicht zusätzlich gegen Veränderung geschützt, können diese sehr leicht abgeändert oder verfälscht werden. Diese zusätzlichen Sicherheitsmaßnahmen können mitunter aufwendig sein und bieten auch keinen hundertprozentigen Schutz. Ältere PDF (Portable Document Format) Dokumente wurden früher standardmäßig mit 40Bit verschlüsselt. Durch die in den letzten Jahren exponentiell gestiegene Rechenleistung handelsüblicher Hardware können diese nun leicht entschlüsselt werden. Des Weiteren kann es auch Probleme geben, wenn eine Veränderung vom Autor selbst gewünscht wird. Durch eine große Anzahl von Versionen von ein und demselben Dokument wird ein z.B. Zitieren schwierig. [Efoo]

2.1.2.3 Anfällig für Beschädigungen

Durch deren Abhängigkeit von Hardware und Software sind digitale Dokumente auch anfällig für Beschädigungen. Durch die hohe Kapazitätsdichte aktueller Datenträger kann schon bei geringer Beschädigung eine große Menge an Daten verloren gehen. Dasselbe kann auch bei den verwendeten Programmen auftreten. Wird ein Dokument bearbeitet und gespeichert, kann erst beim nächsten Öffnen festgestellt werden, ob das Dokument noch in Ordnung ist. Es kann dadurch eine mögliche Beschädigung erst sehr spät festgestellt werden. Aufgrund dessen ist eine gut überlegte Backupstrategie bzw. Langzeitarchivierung obligatorisch. Einen groben Überblick über die Ausfallhäufigkeit verschiedener Speichermedien liefert nachfolgende Grafik, siehe Abbildung 2.5 auf Seite 30. [Efoo]

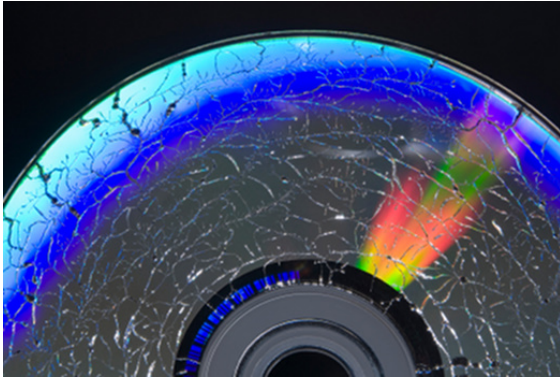


Abbildung 2.3: Beschädigte CD [Sch12]



Abbildung 2.4: Festplattencrash [Dlu12]

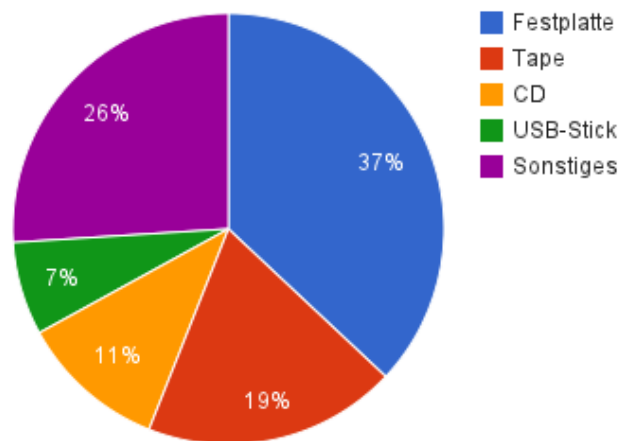


Abbildung 2.5: Speichermedien - Anteil der Vorfälle von Datenverlust © STATISTA

2.1.2.4 Unsicherheit bei Übermittlung

Der Vorteil der schnellen Übermittlung von einem Punkt zu einem anderen birgt auch ein Risiko. Wenn diese Übermittlung über unsichere oder offene Netze erfolgt, so kann das Dokument abgefangen, geändert und diese Fälschung dann weiter gesendet werden. Um dieses Risiko zu minimieren, muss entweder die Verbindung, das Dokument oder beides verschlüsselt werden. [Efoo]

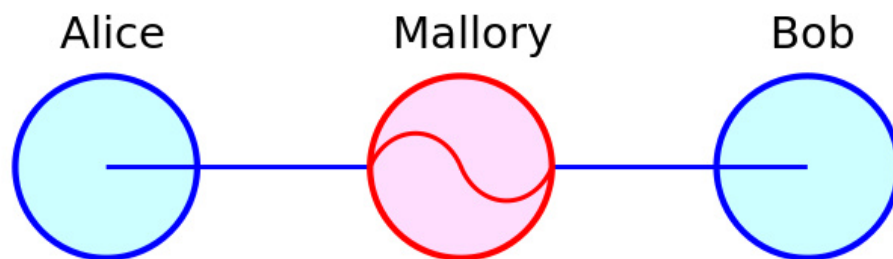


Abbildung 2.6: Man in the middle Attacke © WIKIPEDIA

Eine Lösung für dieses Problem wäre die Nutzung von HTTPS (Hypertext Transfer Protocol Secure). Bei diesem Protokoll wird zwischen Sender und Empfänger eine verschlüsselte Verbindung aufgebaut, das ein Abhören oder ein Abfangen von Nachrichten unmöglich macht.

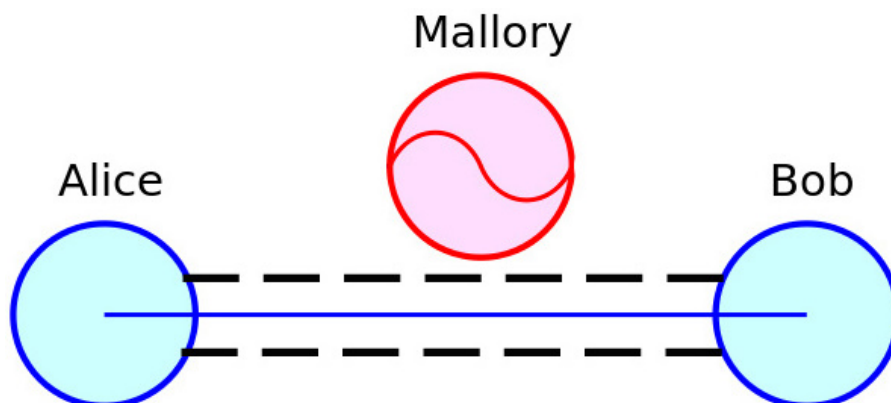


Abbildung 2.7: Verbindung durch HTTPS geschützt

2.1.2.5 Problem der Langzeitarchivierung

Der vielleicht größte Nachteil betrifft die Langzeitarchivierung. Betrachtet man die rasanten Technologiesprünge der letzten Jahre im Bereich Software und Hardware, so wird einem bewusst, wie vergänglich Datenträger und Dateiformate sind. Die Vergangenheit hat gezeigt, dass der Zeitraum, in dem ein Medium primär genutzt wird, ca. 10-15 Jahre beträgt. [Efoo] Am Beispiel der Verkaufszahlen von CD-Rohlingen kann diese Tendenz genau beobachtet werden.

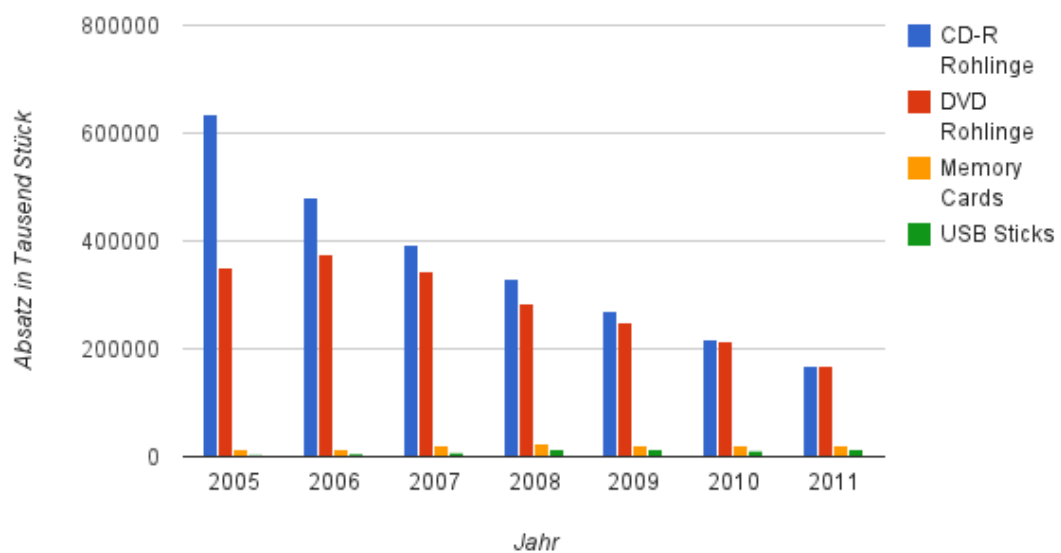


Abbildung 2.8: Absatz von Speichermedien auf dem Konsumentenmarkt in Deutschland von 2005 bis 2011 © STATISTA

Wurden im Jahr 2005 noch ca. 636 Millionen Stück CD-Rohlinge verkauft, so waren das sechs Jahre später nur mehr ca. 168 Millionen Stück. Das ist ein Einbruch von 73,6% innerhalb von sechs Jahren. Die Verkaufszahlen bei Memory Cards und USB Sticks scheinen in der Statistik verschwindend gering zu sein. Das liegt an dem begrenzten Speichervolumen von CD-Rs und DVDs. Eine durchschnittliche Memory Card von 32GB kann die Daten von knapp über 45 CDs speichern. Dieser Vergleich zeigt dass die CD aber auch die DVD als Speichermedium bereits ausgedient hat. In naher Zukunft wird CD und DVD dieselbe Bedeutung haben wie die 3,5" Floppy Diskette in heutiger Zeit hat.

2.2 Dokumentenformate - Voraussetzungen und Anforderungen

Die in digitalen Bibliotheken verwendeten Formate sollten gewisse Eigenschaften aufweisen. Diese Eigenschaften sind notwendig, um auch all die Vorteile nutzen zu können, die digitale Dokumente gegenüber herkömmlichen Dokumenten bieten. Darüber hinaus muss auch für die Langzeitarchivierung vorgesorgt werden. Gerade die Kriterien zur Langzeitarchivierung sollten einen großen Einfluss auf die verwendeten Dokumentenformate haben. Folgende Eigenschaften sollten erfüllt sein.

2.2.1 Offenheit

Das Format sollte offen und für jeden kostenfrei zugänglich sein. Es darf keine Einschränkungen von Organisationen oder Firmen aufweisen. Die Spezifikation des Formats sollte gut dokumentiert und am besten auch als Standard vorliegen. Diese Punkte sind Voraussetzung für eine mögliche Implementierung in anderen Programmen, wenn Originalprogramme nicht mehr existieren sollten. [AF05]

2.2.2 Verbreitung

Ein weiterer wichtiger Aspekt ist die Akzeptanz des Formats in der Informationsgesellschaft. Eine weite Verbreitung sollte auch ein Kriterium sein. Durch diese ist auch gewährleistet, dass viele Programme damit umgehen können. Möglicherweise gibt es dadurch auch verfügbare Konverter, die nutzbar gemacht werden können. Eine weite Verbreitung liefert auch eine gewisse Sicherheit für deren weitere Verwendung. Ein möglicher Übergang in ein anderes Format kann dadurch nur sehr langsam und gut vorbereitet erfolgen. [AF05]

Da eine weite Verbreitung aber nicht unbedingt mit einer offenen Dokumentation und Spezifikation zusammen hängt zeigt das Beispiel von MS Office in den Versionen vor Office 2007. Durch die weite Verbreitung von Microsoft Betriebssystemen und das darauf laufende MS Office wurden die darin verwendeten Dateiformate gewissermaßen als Standard für den Austausch von Officedokumenten verwendet. Die Spezifikation der Formate war lange nicht öffentlich und wurde erst im Jahr 2008 von Microsoft freigegeben. [Mic10]

2.2.3 Komplexität

Selbst wenn das Format offen und gut dokumentiert ist, sollte darauf Wert gelegt werden, dass der Umfang der Spezifikation nicht zu komplex ist. Je aufwendiger ein Format spezifiziert wurde, desto schwieriger wird auch deren weitere Implementierung. [RW07]

Ein gutes Beispiel liefert hier wieder die Firma Microsoft mit ihrem Format OOXML. Dieses wird in allen Microsoft Office Versionen ab der Version 2007 verwendet. Alternative Office Versionen wie LibreOffice, OpenOffice oder KOffice nutzen dagegen das Format ODF (OpenDocument). Beide Formate sind mittlerweile ein offener Standard ODF: ISO/IEC 26300:2006 und OOXML: ISO/IEC 29500-1:2008, ISO/IEC 29500-2:2008, ISO/IEC 29500-3:2008 und ISO/IEC 29500-4:2008. Der gravierende Unterschied liegt in der Seitenanzahl der beiden Spezifikationen. Gerade Komplexität und Umfang der Spezifikation von OOXML waren einer der großen Kritikpunkte bei der Übernahme als Standard. Die Spezifikation von OOXML ist in vier Dokumente unterteilt und besteht zusammengerechnet aus 7358 Seiten. [ISO08] Die Spezifikation von ODF kommt in der Version 1.0 kommt mit 728 Seiten aus. [ISO06]



Abbildung 2.9: Spezifikation von OOXML in ausgedruckter Form [noo12]

2.2.4 Schutzmechanismen

Das Format selbst sollte auch die Möglichkeit bieten verschiedene Ebenen an Sicherheitseinstellungen zu verwenden. Gerade im Bereich von Urheberrecht und DRM (Digital Rights Management) müssen solche Eigenschaften gegeben sein. Es soll möglich sein Dokumente gegen Veränderung zu schützen, um somit die Authentizität und Integrität gewährleisten zu können. Zusätzlich soll auch eine Verschlüsselung möglich sein um eine weitere noch tiefere Sicherheit anzubieten. In Bezug auf die Langzeitarchivierung sollte aber überlegt werden, wieviel Schutz wirklich notwendig ist. Um ein geschütztes Dokument migrieren zu können, ist eine Bearbeitung unumgänglich. Es muss dazu dann das Wissen über Schlüssel und Verschlüsselungsverfahren vorhanden sein. [RW07]

2.2.5 Selbstdokumentation

Mit Selbstdokumentation ist hier die Unterstützung von Metaformaten gemeint. Das Format sollte mindestens eine Variante dieser Beschreibungsmöglichkeit unterstützen. Der Vorteil liegt darin, dass Daten und Metadaten in einem Dokument vorhanden sind. Ist dies nicht der Fall, so müsste man sich auf externe Programme und deren Möglichkeiten verlassen, um Metainformation speichern zu können. Ein Trennen der Metainformation von einem Dokument in ein externes System hebt auch die Fehler- und Ausfallswahrscheinlichkeit. [AF05]

2.2.6 Robustheit

Hier kommt es darauf an, wie anfällig oder stabil das Format auf Fehler reagiert. Was passiert, wenn es auf Bitebene zu Fehlern kommt z.B. durch eine fehlerhafte Speicherung. Ein kleiner Fehler sollte dabei keine große Auswirkung haben. Das Format soll im schlimmsten Fall wenigstens noch Teile der Information wiedergeben können. Ein weiterer Punkt ist die Kompatibilität zu Versionen davor und möglichen Versionen danach. Gehen bei einem Format-Versionswechsel Informationen verloren oder gibt es die Möglichkeit einer Konvertierung oder Importierung in das neue Format?[RW07]

2.2.7 Abhängigkeit

Das Format sollte so wenig Abhängigkeiten wie möglich aufweisen. Abhängigkeiten können spezielle Hard- und Softwareanforderungen sein. Es sollte da-

durch auch unabhängig von speziellen Betriebssystemen oder Anwendungsprogrammen sein. Für die Verwendung des Formats sollten auch keine externen Ressourcen notwendig sein. Eine externe Ressource würde wieder eine weitere Abhängigkeit bedeuten, die im Falle einer Langzeitarchivierung Probleme bereiten könnte. [RW07]

2.3 eBooks

Elektronische Bücher (eBooks) sind von ihrer Darstellung gedruckten Büchern nachempfunden, können aber weit mehr Funktionalität beinhalten, da diese als digitale Dateien zur Verfügung stehen. Bezüglich erweiterter Funktionalität siehe Abschnitt 2.1.1 auf Seite 26. Das erste eBook mit dem Titel »United States Declaration of Independence« wurde bereits 1971 von dem Projekt Gutenberg publiziert. Das Projekt Gutenberg ist eine Organisation, die es sich zum Ziel gesetzt hat, das kulturelle Erbe zu digitalisieren und zu archivieren. Derzeit kann in dieser Bibliothek auf ca. 38.000 freie eBooks zugegriffen werden. [Gut12]

1971	Die erste digitale Bibliothek mit dem Namen Projekt Gutenberg erstellt das erste eBook mit dem Titel »United States Declaration of Independence«
1993	Zahur Klemath Zapata entwickelt die erste Software »Digital book version 1« um eBooks zu lesen. Der erste darin publizierte Titel lautet »Considered as one of the Fine Arts« von <i>Thomas de Quincey</i>
1995	Amazon beginnt mit dem Verkauf von Büchern über das Internet
1998	erste Webseiten beginnen mit dem Verkauf von eBooks z.B. eReader.com
2000	Steven King bietet sein Buch mit dem Titel »Riding the Bullet« ausschließlich in digitaler Form an
2004	Sony entwickelt den ersten eBook Reader mit der neuen Displaytechnologie eInk
2007	Amazon startet den Verkauf des eBook Reader Kindle in Amerika
2009	Amazon verkauft den Kindle 2. Barnes & Noble startet den Verkauf des eBook Reader Nook
2010	Apple verkauft das iPad mit der Applikation iBooks. Google startet das Service Google Books
2011	Amazon verkauft mehr eBooks als gedruckte Bücher im Verhältnis 105:100 [Rap11]

Tabelle 2.1: Historische Entwicklung von eBooks [Sfe11]

2.3.1 Marktentwicklung

Die Geschichte zeigt es und auch die Verkaufszahlen belegen es, die Zukunft der Bücher ist digital. Entscheidend für diese Entwicklung war die Erfindung neuer Displaytechnik und mobiler Hardware wie Tablets und eReader. Letztere ermöglichen dem Nutzer ein Lesen der Bücher ohne auf einen Desktop Computer oder Laptop angewiesen zu sein. Gerade mit dem Erscheinen der ersten eReader erweiterten auch viele Verlage ihr Angebot mit digitalen Varianten ihrer Bücher. Der Absatz von eBooks steigt von Jahr zu Jahr. [Hof12a] Für den deutschen Markt rechnet man bis zum Jahr 2015 mit einer Absatzsteigerung von bis zu 700%.

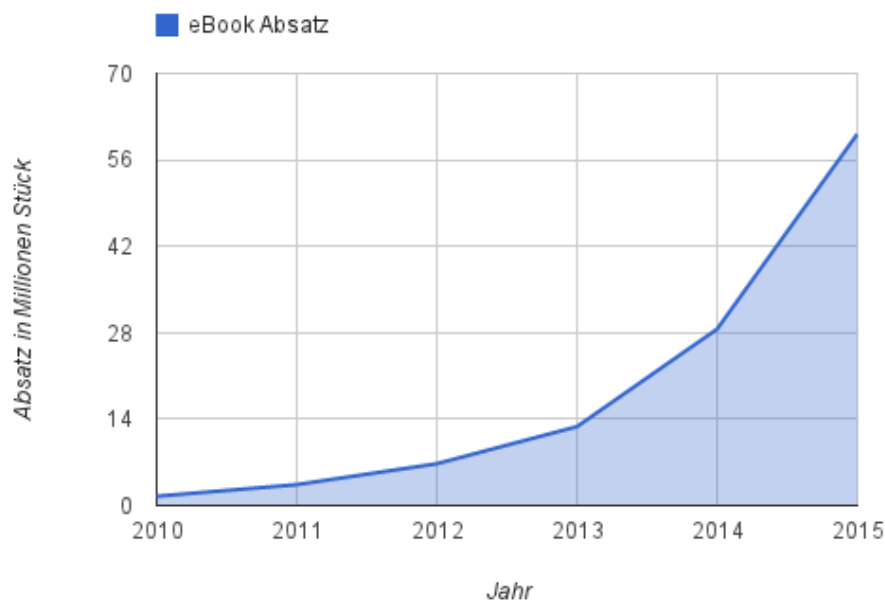


Abbildung 2.10: Prognose zum Absatz von eBooks in Deutschland bis 2015 © STATISTA

2.4 Anzeigeformate

Durch die Vielfalt an eReadern und den dahinter stehenden Firmen sind auch weitere Anzeigeformate entstanden. Das bis dato dominierende Format in digitalen Bibliotheken war und ist PDF (Portable Document Format). Durch die Entwicklung neuer Hardware wurden auch weitere Ansprüche an die Formate

gestellt. Dies führte zu einigen neuen Entwicklungen. Hinter den Formaten stehen teilweise große Unternehmen, die einerseits die Hardware und andererseits die Infrastruktur für den eBook Verkauf anbieten.

Unternehmen	eReader/Tablet	primäres Anzeigeformat
Amazon	Kindle 1-4, Kindle DX, Kindle Fire	Kindle (.azw), Mobipocket (.mobi, .prc)
Apple	iPad 1.Gen - 3.Gen	ePub (.epub)
Barnes & Noble	Nook eReader und Nook Tablet	ePub (.epub)
Sony	PRS 300 - 950SC	ePub (.epub), BBeB (.lrs)

Tabelle 2.2: Zusammenhang: Unternehmen, eReader/Tablet, Anzeigeformat

Die in obiger Tabelle angeführte Hardware an eReadern und Tablets unterstützen auch weiterhin das PDF Format. Es hat auch eine Wandlung des Leseverhaltens stattgefunden. Um eBooks zu lesen war man früher an klobige Hardware und meist spezielle Software gebunden. Das änderte sich mit der breiten Verfügbarkeit von kleinen energiesparenden eReadern.

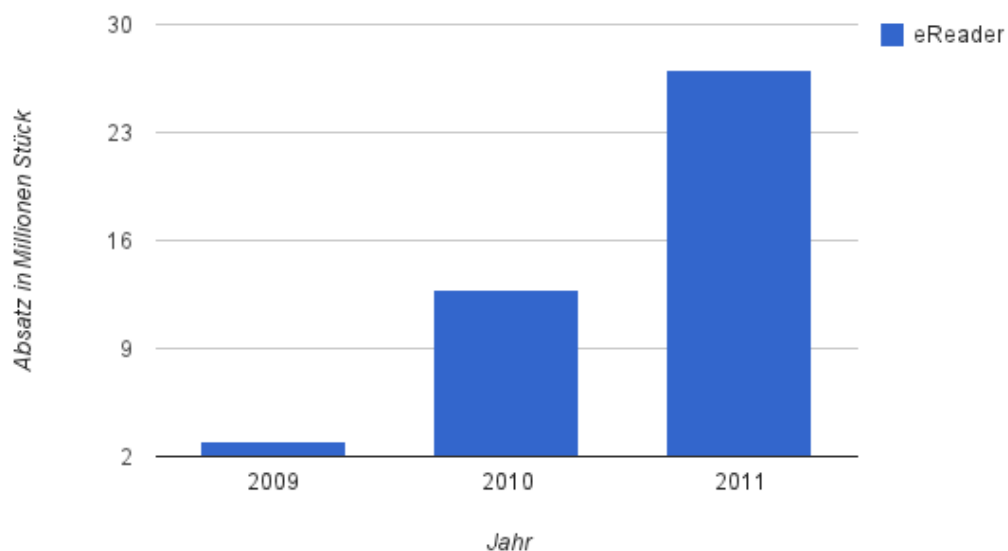


Abbildung 2.11: Absatz von E-Readern weltweit von 2009 bis 2011 © STATISTA

Hier kamen die Vorteile digitaler Dokumente noch besser zur Geltung. Aufgrund dessen wurden von den Herstellern dieser eReader auch eigene Formate entwickelt, von denen sich einige wenige durchgesetzt haben. Diesem anfangs kleinen Teil an eReadern muss in Zukunft auch von Seiten der Verleger und von Seiten digitaler Bibliotheken mehr Beachtung geschenkt werden. Dies belegen die steigenden Verkaufszahlen der letzten Jahre.

Derzeit gibt es ca. 29 Formate, die von eReadern unterstützt werden. Kein eReader kann alle Formate darstellen. Das einzige Format, das alle eReader darstellen können, ist PlainText. Hier zeigt sich auch die Problematik. Anbieter von eBooks müssen überlegen, in welchen Formaten sie ihre Produkte anbieten wollen und gleichzeitig welche Hardware sie mit den Formaten ausschließen. Ein Beispiel für die weite Verbreitung ist das proprietäre Format Kindle (.azw). Dieses Format hat die Verbreitung ausschließlich durch die vom Hersteller Amazon verkauften eReader geschafft. Wie stark der Einfluss eines Herstellers auf Formate sein kann ist durch diese Statistik belegt.

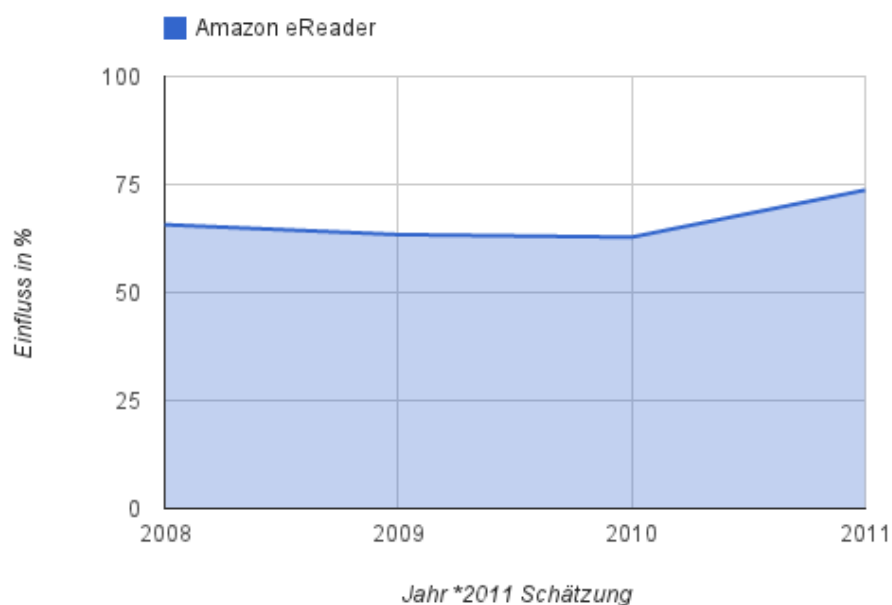


Abbildung 2.12: Marktanteil von Amazon am Absatz von E-Readern weltweit von 2008 bis 2011
© STATISTA

Neuere Versionen dieser Reader unterstützen mittlerweile aber auch andere offene Formate wie ePub. Damit die Käufer eigener Geräte keinen Nachteilen am eBook Markt ausgesetzt sind, bieten Hersteller oft Konverter an, mit denen eine

formatübergreifende Konvertierung möglich ist. Mittlerweile haben sich folgende Formate durch die breite Unterstützung oder durch die bereits breite Verbreitung durchgesetzt.

3 Technologien

Anyone who puts a small gloss on a fundamental technology, calls it proprietary, and then tries to keep others from building on it, is a thief.

(Tim O'Reilly 2000)

3.1 Häufig genutzte Anzeigeformate

3.1.1 PDF

PDF steht für Portable Document Format. Das Format wurde im Jahr 1993 von der Firma Adobe Systems veröffentlicht und dient zur Beschreibung und Darstellung von Dokumenten. Ziel war es, ein Format zu schaffen, das Dokumente erstellt deren Darstellung unabhängig vom Anwendungsprogramm immer gleich aussehen soll. PDF war bis zur Version 1.7 ein proprietäres, aber offenes Dateiformat. Mit der Version 1.7 wurde es standardisiert und ist seither offener Standard ISO 32000-1:2008. Mit diesem Format ist es möglich, Dokumente zu erstellen die unabhängig von Anwendungsprogrammen, Hardware und Betriebssystemen immer gleich dargestellt werden. Diese Eigenschaft war verantwortlich, dass das PDF ein beliebtes und weit verbreitetes Format im Internet wurde. Auch jetzt werden noch viele digitale Dokumente mit diesem Format angeboten. [Leu09]

3.1.1.1 Aufbau eines PDF Dokumentes

Die Erklärung erfolgt hier in zwei Schritten. Es wird in weiterer Folge nur die Dateistruktur und die Dokumentenstruktur genauer betrachtet. Für detailliertere Informationen kann hier der bereits erwähnte Standard ISO 32000-1:2008 herangezogen werden. Die Struktur einer PDF Datei besteht aus vier Teilen.

Header: Dieser beinhaltet die Version der PDF Spezifikation mit der das erstellte Dokument kompatibel ist. Beispiel eines Headers %PDF-1.0 oder %PDF-1.7

Body: Dieser Teil beinhaltet alle Objekte, die zur Darstellung und Inhalt des Dokumentes notwendig sind. PDF kennt acht Basistypen von Objekten wie Boolean, Integer, Strings, Names, Arrays, Dictionaries, Streams, und das Null-Objekt. Objekte im Body wären z.B. Schriftart, Seiten, Bilder, usw.

Cross-reference table: Dieser beinhaltet Verweise auf indirekte Objekte, die verwendet oder benötigt werden. Indirekte Objekte sind im gesamten Dokument referenzierbar z.B. Überschriften oder Bilder.

Trailer: Dieser gibt an, wo sich der Cross-reference table im Dokument befindet. Will man auf diesen zugreifen, wird das Ende der Datei ausgelesen. Bei einer PDF Datei wird also immer zuerst das Ende gelesen um schnell an relevante Information zu kommen.[Adoo8]

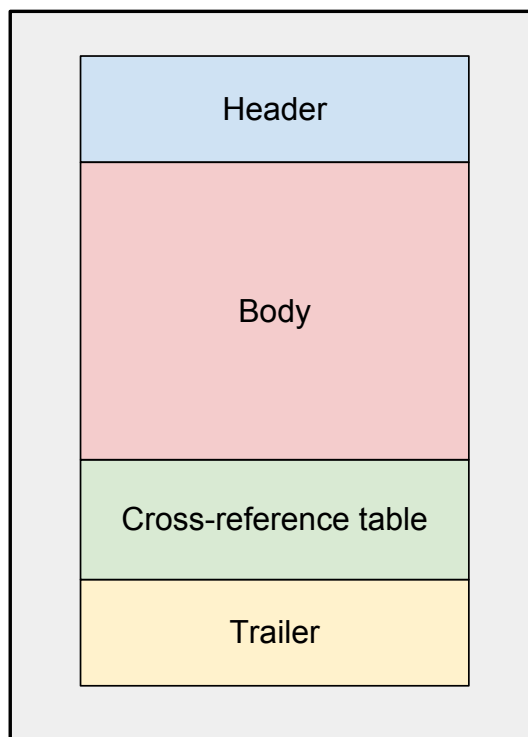


Abbildung 3.1: Dateistruktur einer PDF Datei [Adoo8]

Die Dokumentenstruktur (siehe Abbildung 3.2 auf Seite 43) ist streng hierarchisch aufgebaut und legt fest, wie die einzelnen Objekte verwendet werden. Das Dokument wird hierbei nach und nach durch die notwendigen Objekte zusammengebaut. Es wird auch festgelegt, wie das Dokument am Bildschirm dargestellt wird. Das Wurzelement ist hierbei der Dokumentkatalog. Aufbauend auf diesen werden viele Eigenschaften des PDF Dokumentes festgelegt. Eigenschaften können sein: der Typ, Anzahl der Seiten, das Layout des Dokumentes, was beim Öffnen passieren soll, usw. Der Dokumentkatalog kann derzeit 29 Eigenschaften beinhalten, von denen derzeit nur zwei obligatorisch sind (Typ und Pages). [Adoo8]

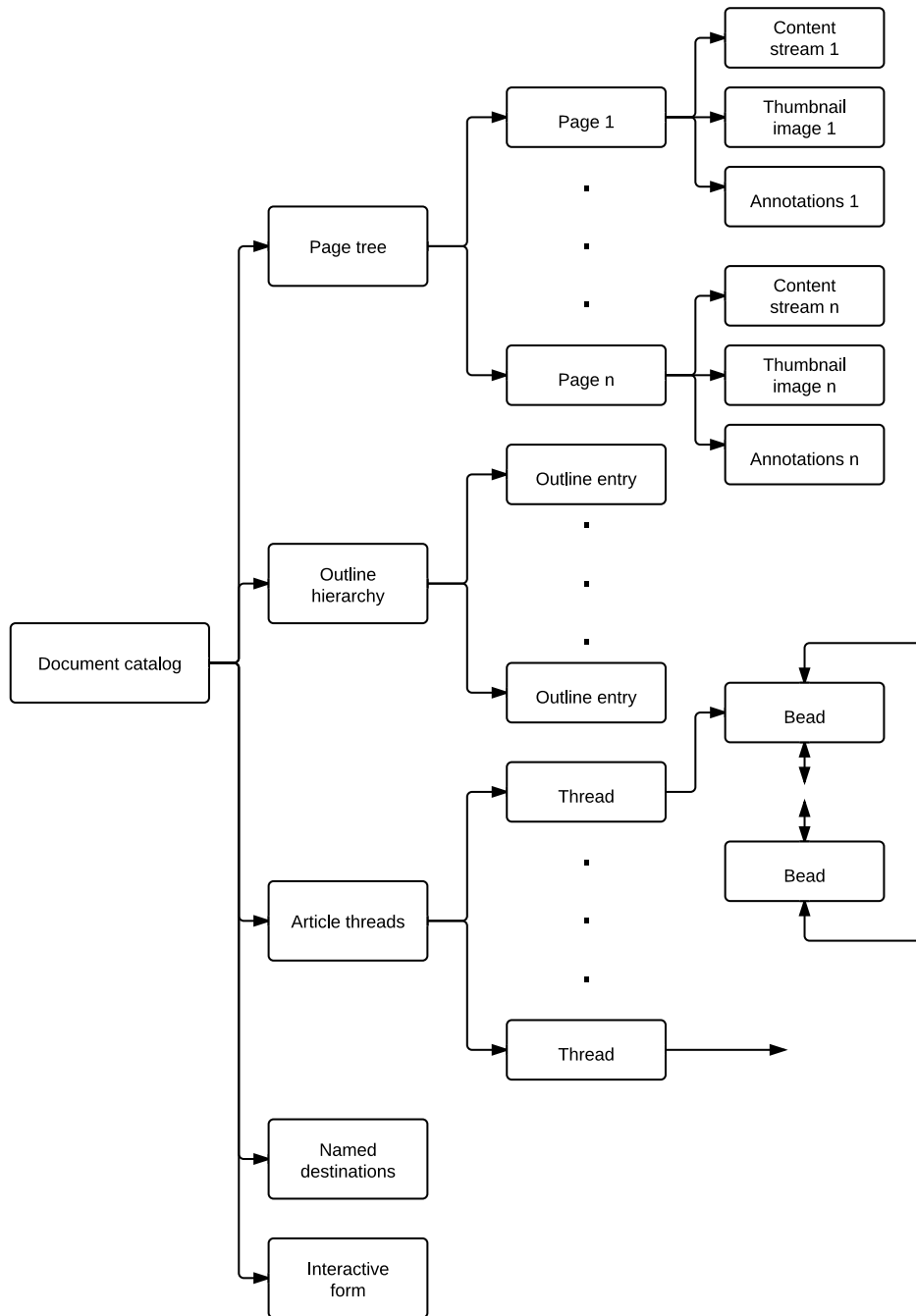


Abbildung 3.2: Aufbau eines PDF Dokumentes [Adoo8]

3.1.1.2 Vorteile von PDF [Ran07]

- Es ist plattformübergreifend. PDF-Dateien können auf jedem System mit jeder kompatiblen Software geöffnet werden und sehen immer gleich aus, egal ob das Betriebssystem auf Windows, Mac, Unix/Linux, Solaris, iOS, Android, oder anderen basiert. Durch diese Eigenschaft ist es einfach Dokumente mit anderen Nutzern zu teilen.
- Dokumente in diesem Format sind im Vergleich zu anderen Office Formaten sehr kompakt und verhältnismäßig klein. Grund dafür ist die Optimierung, die passiert, wenn ein Dokument in ein PDF Format konvertiert wird. Des weiteren können auch unterschiedliche Quelldateien bestehend aus z.B. Bildern, Text oder Tabellen zu einem Dokument zusammengefasst werden.
- Durch die Standardisierung ist auch viel unterschiedliche Software entstanden, die mit dem PDF Format umgehen kann. Es gibt somit eine breite Unterstützung, um Dokumente in das PDF Format zu konvertieren. Durch die umfangreiche Technologie hinter PDF lassen sich fast alle Dokumente konvertieren.
- Das Format unterstützt gewisse Sicherheitsoptionen. So ist es möglich, das Drucken und sogar das Öffnen selbst zu unterbinden. Wird ein Dokument erstellt, kann der Autor aus diversen Sicherheitsoptionen wählen.
- Es gibt eine gute und umfangreiche Unterstützung von anderen Officeprodukten wie MS Office oder LibreOffice. In diesen Umgebungen ist ein PDF Konverter integriert. Für andere Varianten gibt es die Möglichkeit einen »PDF-Drucker« zu verwenden. Mit diesem ist es möglich, die Ausgabe nicht auf einen Drucker im herkömmlichen Sinn auszugeben, sondern den Druck in eine Datei umzuleiten. Dadurch gilt: Alles was druckbar ist, kann auch in ein PDF gespeichert werden.
- Die Programme, um PDF-Dateien zu lesen, sind zahlreich und kostenlos. Viele Browser können bereits PDFs darstellen ohne ein Plugin verwenden zu müssen.
- Nahezu jeder eReader und jede Softwareplattform kann PDF-Dateien lesen und darstellen. Obwohl es mittlerweile weitere etablierte Formate am Markt gibt, wird PDF noch immer von einer sehr breiten Masse an Hardware wie z.B. Smartphones unterstützt.

3.1.1.3 Nachteile von PDF [Leu09]

- Es gibt mittlerweile viele unterschiedliche Versionen. Adobe hat in den letzten Jahren ca. alle zwei Jahre eine neue Version veröffentlicht. Jede neue Version unterstützt neue Features. Diese können von älteren PDF-Readern nicht unterstützt werden. Man ist also immer auf aktuelle Updates seiner PDF-Reader Software angewiesen. Das könnte wiederum die »Plattformunabhängigkeit« etwas einschränken. Die Vergangenheit hat gezeigt, dass neue PDF-Reader Versionen als erstes auf Windows Plattformen veröffentlicht werden. Würde nun ein neues Dokument mit neuen Features auf einer Windows Plattform erstellt, so könnte es Probleme geben diese auf einer z.B. Solaris Plattform zu öffnen, da Solaris diese neuen Features noch nicht in deren Version des PDF-Reader unterstützt.
- PDF versucht alles für jedermann zur Verfügung zu stellen. Das hat zur Folge, dass PDF für spezielle Aufgaben nicht so effizient eingesetzt werden kann wie z.B. ein darauf optimiertes Format.
- In einigen Fällen kann ein HTML (Hypertext Markup Language) Dokument effizienter eingesetzt werden. Wird z.B. ein großflächig und komplex angelegtes Layout verwendet, könnte es Probleme mit PDF geben. Weiters kann auch das Suchen in PDF Dokumenten durch Unterscheidungsschwierigkeiten einzelner Buchstaben eingeschränkt sein.
- PDF Dokumente wurden nicht zum nachträglichen Bearbeiten vorgesehen. Sie sind nur durch zusätzliche, meist kostenpflichtige Software editierbar. Dieses Bearbeiten ist aber sehr eingeschränkt. Kleinere Änderungen sind möglich, umfangreiches Hinzufügen von Bildern und Text ist nicht möglich. Diese Einschränkungen sind auch mit aktueller Software nicht zu umgehen.
- Es ist sehr schwer und aufwendig Daten aus einem PDF zu exportieren. Während es für viele andere Dokumentenformate Konverter gibt, um ein PDF zu erzeugen, ist es umgekehrt nur eingeschränkt bis gar nicht möglich.
- PDF Dokumente haben ein fixes Layout. Sie können sich nicht unterschiedlichen Lesegeräten anpassen. In Zeiten von Smartphones, Tablets und größeren Computerbildschirmen ist dies wohl der Größte aller Nachteile. Reicht die Auflösung oder Größe des Displays nicht aus, so muss aufwendig gescrollt oder gezoomt werden, um alle Inhalte des Dokumentes erfassen zu können.

3.1.2 EPUB

EPUB ein Akronym für electronic publication ist ein freier und offener eBook Standard vom International Digital Publishing Forum (IDPF). EPUB basiert auf einer Reihe freier Standards wie XML (Extensible Markup Language), XHTML (Extensible Hypertext Markup Language), CSS (Cascading Style Sheets), NCX (Navigation Center eXtended), Dublin Core und Zip. EPUB hat im Gegensatz zu PDF ein flexibles Layout und kann sich damit auch Displays unterschiedlicher Größe leicht anpassen. Mittlerweile ist EPUB in der Version 3.0 verfügbar. Die Spezifikation von EPUB 3 ist in vier Teile gegliedert, die jeweils unterschiedliche Bereiche definieren. [IDP12a]

3.1.2.1 EPUB Publications 3.0

Diese definiert die Semantik und übergreifende Konformitätsanforderungen für EPUB-Publikationen. Weiters beschreibt dieses Dokument den Zusammenhang weiterer Ressourcen, um eine EPUB konforme Publikation zu erstellen. Ein weiterer Bereich sind auch Metadaten und deren Unterstützung. [IDP12e]

3.1.2.2 EPUB Content Documents 3.0

Dieser Teil definiert die Verwendung von HTML5, SVG und CSS im Bereich der EPUB Publikation. Es werden, wenn nicht anders angegeben, Beschreibung der Definitionen, der Semantik, Struktur und Verarbeitung aus den Grundspezifikationen übernommen. (z.B. HTML5) [IDP12b]

3.1.2.3 EPUB Open Container Format (OCF) 3.0

Diese Spezifikation beschreibt die Struktur des Formats. Das Format selbst besteht aus mehreren Dateien, die spezielle Aufgaben haben und in einer definierten Struktur gespeichert sind. Das OCF beschreibt, welche Dateien notwendig sind, um ein EPUB Format sein zu können. Diese werden in einer einzelnen Containerdatei zusammengefasst. Die Containerdatei bietet Vorteile beim Tausch oder beim Transport und bietet einen leichten Zugang zu den Einzeldateien. [IDP12d]

3.1.2.4 EPUB Media Overlays 3.0

Dieser Teil der Spezifikation beschäftigt sich mit der Beschreibung und Verwendung von SMIL (Synchronized Multimedia Integration Language). Weiters wird auch die Verwendung von Audio und Video in Verbindung mit dem EPUB Format definiert.[IDP12c]

In Summe beinhaltet die Spezifikation von EPUB schon viele andere offene Standards. Eine Übersicht der Implementierungsmöglichkeiten und Technologien soll diese Grafik zeigen.

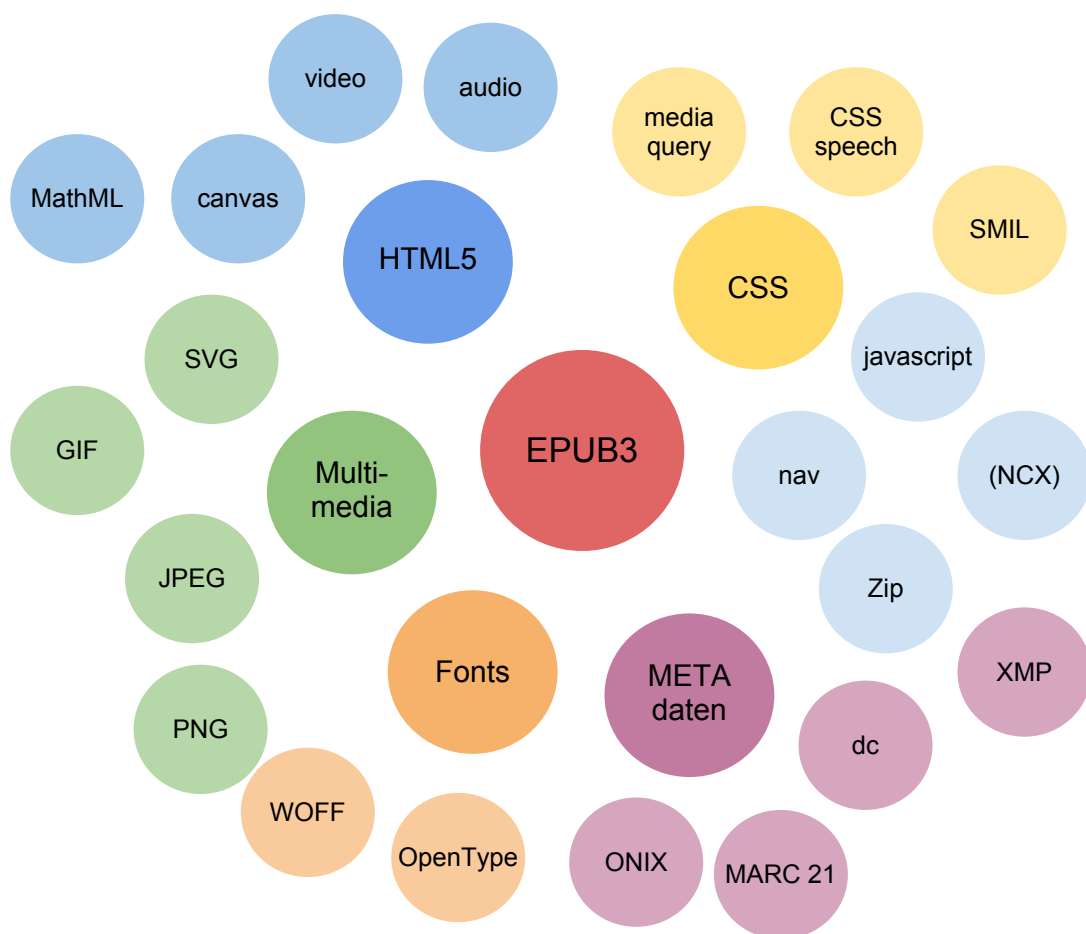


Abbildung 3.3: Mögliche Technologien in EPUB 3 (angelehnt an [MW11])

3.1.2.5 Aufbau eines EPUB Dokumentes

Ein EPUB Dokument ist im Grunde ein Archiv und besteht aus mehreren Dateien und Ordnern. Den Aufbau einer solchen Datei soll diese Grafik veranschaulichen. All das, was in diesem Archiv vorkommen kann, ist grün hinterlegt, die Ordner selbst sind gelb hinterlegt.

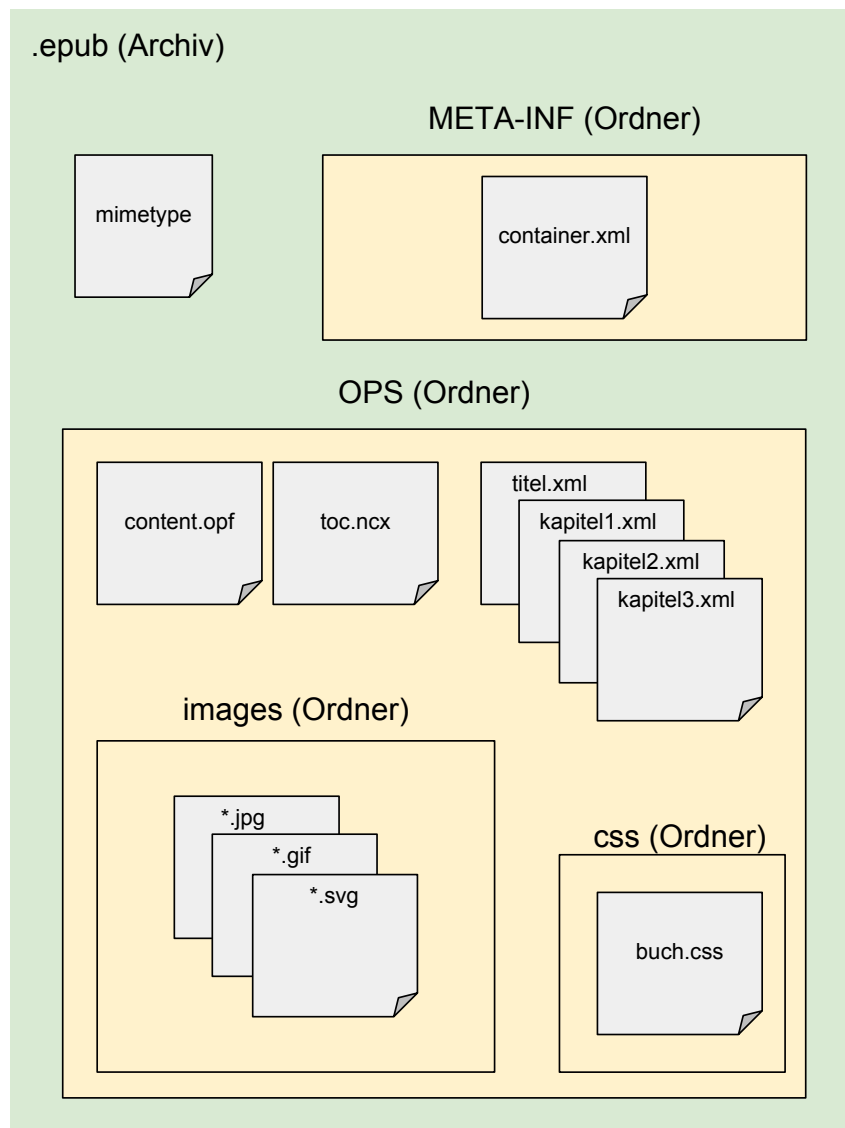


Abbildung 3.4: Aufbau eines EPUB Dokumentes

mimetype: Diese Datei sagt dem Reader oder der Software, welcher Mimetype hier zu erwarten ist. In dem Fall wäre der Mimetype »application/epub+zip«

META-INF Ordner: In diesem Ordner muss mindestens die Datei container.xml vorhanden sein. Diese liefert dem Reader weitere Informationen und Verweise. Weiters können in dem Ordner Informationen zu Verschlüsselung, Metadaten, Rechtemanagement und digitalen Signaturen sein. [IDP12d]

OPS Ordner: In diesem Ordner können sich noch weitere Unterordner wie z.B. »images« und »css« finden. Im Ordner selbst befinden sich der Inhalt des Buches aufgeteilt in weitere Quelldateien (*.xml,*.xhtml), die *.ncx Datei zum Aufbau des Inhaltsverzeichnis und eine Datei mit Endung *.opf. Auf diese wird vom container.xml im Ordner META-INF referenziert. Die *.odf Datei ist unterteilt in mehrere Bereiche wie z.B. Metadaten, Css, Bilder und weiterer Information über den allgemeinen Aufbau des Buches. Die Struktur wird im nachfolgenden Beispiel noch genauer beschrieben.

3.1.2.6 Beispiel eines EPUB Dokumentes

Der Aufbau eines EPUB Dokumentes soll an einem realen Beispiel gezeigt werden. Dazu wurde das frei downloadbare eBook mit dem Titel »The blonde Lady« von der Seite epubbooks.com lokal gespeichert. Die Datei mit der Endung *.epub kann mit jedem Packprogramm wie Winzip, WinRAR oder 7-Zip geöffnet werden.

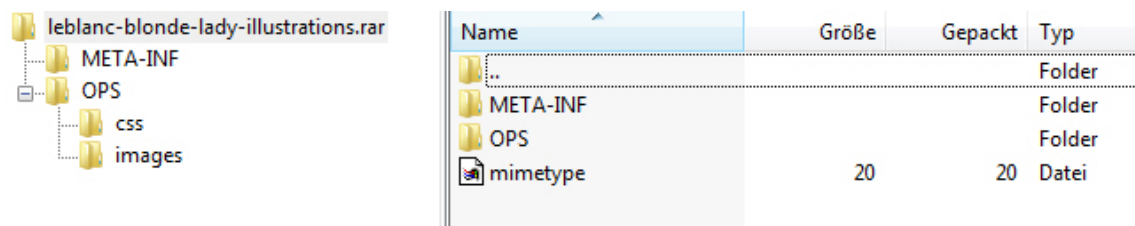


Abbildung 3.5: EPUB Ordnerstruktur

Name	Pfad
OPS	
META-INF	
images	OPS\
css	OPS\
mimetype	
container.xml	META-INF\
title.xml	OPS\
footnotes.xml	OPS\
epubbooksinfo.xml	OPS\
epb.opf	OPS\
epb.ncx	OPS\
chapter-008.xml	OPS\
chapter-007.xml	OPS\
chapter-006.xml	OPS\
chapter-005.xml	OPS\
chapter-004.xml	OPS\
chapter-003.xml	OPS\
chapter-002.xml	OPS\
chapter-001.xml	OPS\
book-02.xml	OPS\
book-01.xml	OPS\
titlepage.css	OPS\css\
book.css	OPS\css\
frontis.jpg	OPS\images\
epubbooks-logo.png	OPS\images\
circle.jpg	OPS\images\
bookcover.jpg	OPS\images\

Abbildung 3.6: Auflistung aller in dem Buch vorhandenen Dateien und Ordner

In weiterer Folge werden die wichtigsten Dateien näher beschrieben. Das im Ordner META-INF abgelegte container.xml referenziert dabei auf das Wurzelement im OPS Ordner. In diesem Fall auf die Datei epb.ops.

```
<?xml version="1.0" encoding="UTF-8"?>
<container xmlns="urn:oasis:names:tc:opendocument:xmlns:container" version="1.0">
  <rootfiles>
    <rootfile full-path="OPS/epb.opf" media-type="application/oebps-package+xml"/>
  </rootfiles>
</container>
```

Quellcode 3.1: META-INF/container.xml

Die referenzierte epb.ops Datei ist der Mittelpunkt des gesamten Buches. In dieser findet sich Information über das Buch selbst (Metainformation) und referenziert auf jede einzelne Inhaltsdatei (chapter-001.xml, chapter-002.xml) auf die im Buch verwendeten CSS Styles und verwendeten Bilder. Des weiteren verweist sie auf die epb.ncx Datei, die für den Aufbau des Inhaltsverzeichnis notwendig ist. Diese Beispieldatei entspricht, wie in Codezeile zwei ersichtlich, noch der package-version 2.0 und nicht der aktuellen package-version 3.0.[IDP12e]

```
<?xml version="1.0" encoding="UTF-8"?>
<package xmlns="http://www.idpf.org/2007/opf" unique-identifier="EPB-UUID" version="2.0">
  <metadata xmlns:opf="http://www.idpf.org/2007/opf"
    xmlns:dc="http://purl.org/dc/elements/1.1/">
    <dc:title>The Blonde Lady</dc:title>
    <dc:creator opf:role="aut" opf:file-as="Leblanc, Maurice">Maurice Leblanc</dc:creator>
    <dc:creator opf:role="ill" opf:file-as="Boehm, H. Richard">H. Richard Boehm</dc:creator>
    <dc:date opf:event="original-publication">1910</dc:date>
    <dc:publisher>epubBooks (www.epubbooks.com)</dc:publisher>
    <dc:date opf:event="epub-publication">2011-01-10</dc:date>
    <dc:subject>Detective</dc:subject>
    <dc:subject>Adventure</dc:subject>
    <dc:subject>Mystery</dc:subject>
    <dc:source>Project Gutenberg</dc:source>
    <dc:rights>
      Provided for free by epubBooks.com. Not for commercial use.
      This EPUB eBook is released under a Creative Commons (BY-NC-ND/3.0) Licence.
      Source text and images are in the Public Domain.
    </dc:rights>
    <dc:identifier id="EPB-UUID">urn:uuid:A1989627-6C07-1014-9E83-850FB41434CC</dc:
      identifier>
    <dc:language>en-gb</dc:language>
  </metadata>
  <manifest>
    <!-- Content Documents -->
    <item id="titlepage" href="title.xml" media-type="application/xhtml+xml"/>
    <item id="epubbooksinfo" href="epubbooksinfo.xml" media-type="application/xhtml+xml"/>
    <item id="book-01" href="book-01.xml" media-type="application/xhtml+xml"/>
    <item id="chapter-001" href="chapter-001.xml" media-type="application/xhtml+xml"/>
    <item id="chapter-002" href="chapter-002.xml" media-type="application/xhtml+xml"/>
    <item id="chapter-003" href="chapter-003.xml" media-type="application/xhtml+xml"/>
    <item id="chapter-004" href="chapter-004.xml" media-type="application/xhtml+xml"/>
    <item id="chapter-005" href="chapter-005.xml" media-type="application/xhtml+xml"/>
    <item id="chapter-006" href="chapter-006.xml" media-type="application/xhtml+xml"/>
    <item id="book-02" href="book-02.xml" media-type="application/xhtml+xml"/>
    <item id="chapter-007" href="chapter-007.xml" media-type="application/xhtml+xml"/>
    <item id="chapter-008" href="chapter-008.xml" media-type="application/xhtml+xml"/>
    <item id="footnotes" href="footnotes.xml" media-type="application/xhtml+xml"/>

    <!-- CSS Style Sheets -->
    <item id="title-page-css" href="css/titlepage.css" media-type="text/css"/>
    <item id="main-css" href="css/book.css" media-type="text/css"/>

    <!-- Images -->
    <item id="epubbooks-logo" href="images/epubbooks-logo.png" media-type="image/png"/>
    <item id="img-bookcover-jpeg" href="images/bookcover.jpg" media-type="image/jpeg"/>
    <item id="img-frontis-jpeg" href="images/frontis.jpg" media-type="image/jpeg"/>

    <!-- NCX -->
    <item id="ncx" href="epb.ncx" media-type="application/xhtml+xml"/>
  </manifest>
```

```

<spine toc="ncx">
  <itemref idref="titlepage" linear="yes"/>
  <itemref idref="epubbooksinfo" linear="yes"/>
  <itemref idref="book-01" linear="yes"/>
  <itemref idref="chapter-001" linear="yes"/>
  <itemref idref="chapter-002" linear="yes"/>
  <itemref idref="chapter-003" linear="yes"/>
  <itemref idref="chapter-004" linear="yes"/>
  <itemref idref="chapter-005" linear="yes"/>
  <itemref idref="chapter-006" linear="yes"/>
  <itemref idref="book-02" linear="yes"/>
  <itemref idref="chapter-007" linear="yes"/>
  <itemref idref="chapter-008" linear="yes"/>
  <itemref idref="footnotes" linear="no"/>
</spine>
</package>

```

Quellcode 3.2: OPS/epb.opf

Weitere Teile in der epb.opf Datei sind der Abschnitt <manifest> und der Abschnitt <spine>. Im Bereich Manifest werden alle Dateien aus dem das Buch besteht referenziert. Der Bereich Spine definiert, in welcher Reihenfolge die einzelnen Inhalte dem Leser angeboten werden z.B. in der Vorwärts- und Rückwärtsnavigation eines eReaders. [IDP12e]

Die *.ncx Datei definiert die Navigationsstruktur des Dokumentes. Diese Datei ist die Grundlage für das klickbare Inhaltsverzeichnis in der Leseumgebung. In der Spezifikation von EPUB 3 wird die Funktion der *.ncx Datei durch die Eigenschaft »nav« ersetzt. Um aber eine Abwärtskompatibilität mit eReadern zu gewährleisten, die noch nach der EPUP Version 2.0 spezifiziert sind, wird die *.ncx Datei noch sehr oft zusätzlich mitgeführt. Lesegeräte, die nach der Version 3.0 spezifiziert sind, ignorieren hingegen die *.ncx Datei und greifen auf die Eigenschaft »nav« zu.[IDP12e]

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ncx
  PUBLIC "-//NISO//DTD ncx 2005-1//EN" "http://www.daisy.org/z3986/2005/ncx-2005-1.dtd">
<ncx xmlns="http://www.daisy.org/z3986/2005/ncx/" version="2005-1">
  <head>
    <!--The following four metadata items are required for all
          NCX documents, including those conforming to the relaxed
          constraints of OPS 2.0-->
    <meta name="dtb:uid" content="A1989627-6C07-1014-9E83-850FB41434CC"/>
    <meta name="epub-creator" content="epubBooks (www.epubbooks.com)"/>
    <meta name="dtb:depth" content="2"/>
    <meta name="dtb:totalPageCount" content="0"/>
    <meta name="dtb:maxPageNumber" content="0"/>
  </head>
  <docTitle>
    <text>The Blonde Lady</text>
  </docTitle>
  <docAuthor>
    <text>Maurice Leblanc</text>
  </docAuthor>
  <navMap>
    <navPoint id="navpoint-1" playOrder="1">

```

```

    <navLabel>
      <text>Title Page</text>
    </navLabel>
    <content src="title.xml"/>
  </navPoint>
  <navPoint id="navpoint-2" playOrder="2">
    <navLabel>
      <text>epubBooks Information</text>
    </navLabel>
    <content src="epubbooksinfo.xml"/>
  </navPoint>
  <navPoint id="navpoint-3" playOrder="3">
    <navLabel>
      <text>FIRST EPISODE</text>
    </navLabel>
    <content src="book-01.xml"/>
    <navPoint id="navpoint-4" playOrder="4">
      <navLabel>
        <text>CHAPTER I</text>
      </navLabel>
      <content src="chapter-001.xml"/>
    </navPoint>
    <navPoint id="navpoint-5" playOrder="5">
      <navLabel>
        <text>CHAPTER II</text>
      </navLabel>
      <content src="chapter-002.xml"/>
    </navPoint>
    . . . . .
    <navPoint id="navpoint-9" playOrder="9">
      <navLabel>
        <text>CHAPTER VI</text>
      </navLabel>
      <content src="chapter-006.xml"/>
    </navPoint>
  </navPoint>
  <navPoint id="navpoint-10" playOrder="10">
    <navLabel>
      <text>SECOND EPISODE</text>
    </navLabel>
    <content src="book-02.xml"/>
    <navPoint id="navpoint-11" playOrder="11">
      <navLabel>
        <text>CHAPTER I</text>
      </navLabel>
      <content src="chapter-007.xml"/>
    </navPoint>
    <navPoint id="navpoint-12" playOrder="12">
      <navLabel>
        <text>CHAPTER II</text>
      </navLabel>
      <content src="chapter-008.xml"/>
    </navPoint>
  </navPoint>
  <navPoint id="navpoint-13" playOrder="13">
    <navLabel>
      <text>Footnotes</text>
    </navLabel>
    <content src="footnotes.xml"/>
  </navPoint>
</navMap>
</ncx>

```

Quellcode 3.3: OPS/epb.ncx

Der Aufbau eines EPUB Dokumentes scheint im ersten Augenblick etwas kompliziert und aufwändig. Die Struktur ist aber klar und übersichtlich und mit Hilfe von etwas HTML und CSS können EPUB Publikationen leicht editiert und angepasst werden. Als Beispiel wird hier das Logo der TU-Graz in die allgemeine eBook Information hinzugefügt. Dazu sind folgende Schritte notwendig:

1. Das TU-Graz Logo in den »images« Ordner kopieren
2. Hinzufügen einer CSS Styledefinition in die Datei »book.css«

```
#tugraz-logo { margin: 1.5em 0; padding: 0; text-align: center; border: 0; }
```

Quellcode 3.4: CSS Quellcode zur EPUB Erweiterung

3. Erweiterung der Datei »epubbooksinfo.xml« mit folgendem Code.

```
<div id="tugraz-logo">
  <p>Tu-Graz Logo</p>
  
</div>
```

Quellcode 3.5: HTML Quellcode zur EPUB Erweiterung

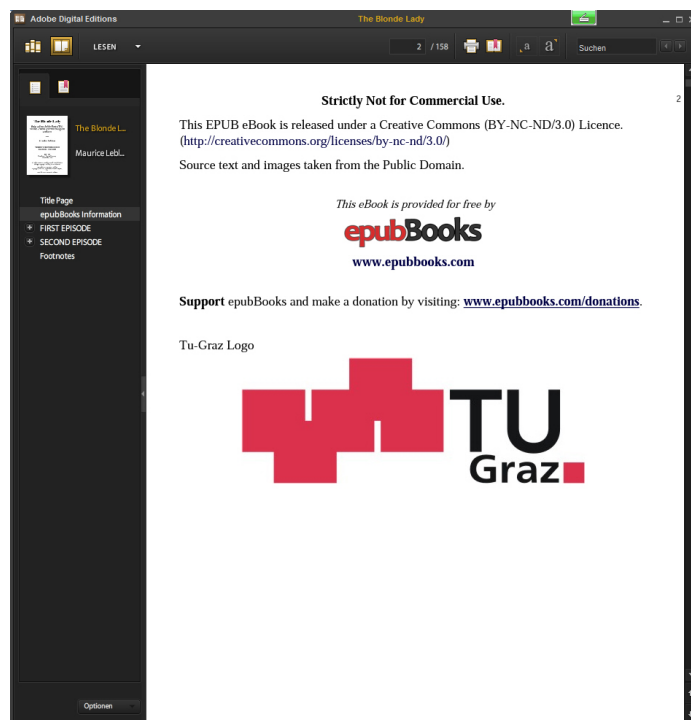


Abbildung 3.7: Editierte EPUB Publikation

3.1.2.7 Vorteile von EPUB [Pra12]

- Das EPUB Format ist ein offener und freier Standard. Es kann damit einfach in andere Software integriert und verwendet werden. Dieser Umstand trägt unter anderem zur breiten Unterstützung bei.
- Das Format unterstützt fließenden Text und kann sich somit jeglicher Displaygröße anpassen. Es können aber auch fixe Auflösungen vorgegeben werden, wo das Layout ein wesentlicher Bestandteil des Lesevergnügens ist (z.B. Kinderbücher)
- Das Format genießt eine breite Unterstützung, wie die Mitgliederliste vom International Digital Publishing Forum zeigt. [IDP12f] Unter den Unterstützern finden sich große Namen wie Google, Apple und Barnes & Noble.
- EPUB unterstützt DRM Funktionalität. Das Format definiert aber nicht deren Implementierung. Dadurch gibt es derzeit zwei große DRM Implementierungen für EPUB.

3.1.2.8 Nachteile von EPUB [Pra12]

- Die Erstellung eines eBooks im EPUB Format ist aufwendig. Es gibt strenge Regeln zum Aufbau der Archivstruktur. Es wird einiges an Zusatzwissen einzelner Technologien wie XML benötigt. Dieser Nachteil bezieht sich aber hauptsächlich auf Autoren und weniger auf Nutzer.
- Die Implementierung mehrerer nicht miteinander kompatibler DRM Systeme. Derzeit weit verbreitet ist ADEPT DRM von Adobe und eine Implementierung von Apple.

3.1.3 Mobipocket / AZW / KF8

Mobipocket ist ein weiterer offener Standard und basiert auf dem International Digital Publishing Forum (IDPF). Das Format wurde ursprünglich von einer französischen Firma seit dem Jahr 2000 entwickelt. Es kann die Dateierweiterung *.prc oder *.mobi haben. Mit Mobipocket lassen sich komplexe Inhalte darstellen. Mobipocket liefert weiters einen hohen Grad an Kompression. Das Format selbst unterstützt Anmerkungen, Korrekturen und weitere interaktive Aktionen des Nutzers. Es unterstützt Fließtext sowie die Darstellung von fixem Layout. [Sai12]

Amazon hat im Jahr 2005 diese Firma übernommen um das für die Kindle Serie verwendete Format *.azw zu entwickeln. [Mob12] AZW ist eine Abwandlung von Mobipocket und proprietär. Es besitzt eine noch bessere Kompression und eine andere DRM Technik. EBooks mit der Endung *.azw werden zu einem sehr großen Teil nur von Amazon Readern (Kindle) verwendet. Durch den DRM Schutz sind solche Bücher auf die ID des Gerätes gebunden. [Sai12]

Mit dem Verkaufsstart von Amazon Kindle Fire wurde ein weiteres Format entwickelt, das sogenannte KF8 (Kindle Format 8) mit der Dateierdung *.azw3. Dieses soll die vorhin genannten Formate mit der Endung *.mobi und *.azw ablösen. Es unterstützt eine Vielzahl von Formatierungsmöglichkeiten in Verbindung mit HTML5 und CSS3. KF8 soll nicht nur für Tablets sondern auch für neuere e-Ink Geräte ausgeliefert werden. [Ama12b]

3.1.4 Apple / iBooks Author

Was Amazon bei den eBook Readern, ist Apple bei den Tablets, Marktführer. So kommt es, dass auch Apple Anfang des Jahres 2012 ein Programm vorgestellt hat, mit welchem man eBooks erstellen mit der Applikation iBooks verwenden und im iTunes-Store verkaufen kann. Bei der Vorstellung von iBooks 2010 wurde noch versichert, ausschließlich auf den offenen EPUB Standard zu setzen. Leider wurde auch hier das EPUB Format so angepasst und abgeändert, dass es nicht mehr dem Standard entspricht. [Gol12]

Apple verwendet nicht standardisierte Namensräume in XML und CSS. Weiters wird auch ein anderer Mimetype übergeben.

```
EPUB: application/epub+zip
Apple: application/x-ibooks+zip
```

Zusätzlich dazu verwendet Apple proprietäre Präfixe in CSS z.B.:

```
-ibooks-layout-hint: anchor page shape;
-ibooks-underline-width: 1.0000px;
```

Das Ergebnis dieser Vorgehensweise ist, dass im iBooks Author erstellte Bücher ausschließlich auf Apple Hardware wie iPads zu lesen sind. EReader, die nach EPUB spezifiziert sind, können diese Bücher nicht lesen. Es gibt derzeit auch keine Möglichkeit, die im iBooks Author erstellten Dokumente zu konvertieren. Fraglich ist, ob Apple hier durch Updates die Funktionalität noch erweitert oder ob sie analog zu Amazon auch den Vorteil in einem geschlossenen System sehen. [Gla12]

3.2 eBook Bundles

Um eine möglichst große Nutzerschicht zu erreichen und weil einige Hersteller auf ihre eigenen proprietären Formate festhalten sind Verleger und Anbieter von eBooks gezwungen, mehrere Formate anzubieten. Der Nutzer hat dadurch die Möglichkeit ein Buch zu kaufen und dieses je nach Lesegewohnheit in meistens drei Formaten (PDF, EPUB, MOBI) herunterzuladen. Mit diesen drei Formaten lassen sich derzeit 100% aller eReader und Tabletutzer erreichen.

Der auf IT-Themen spezialisierte Verlag O'Reilly bietet schon jetzt viele seiner Bücher in mehreren Formaten an. Die Zahl der veröffentlichten Downloads zeigt auch hier, wie wichtig es ist, auf die veränderten Lesegewohnheiten der Nutzer einzugehen. Die Statistik zeigt eine deutliche Tendenz in Richtung mobiler Lesegeräte und deren Formate. O'Reilly verzichtet in all seinen Formaten auf DRM. Dieser sehr positive Umstand bietet dem Nutzer die Möglichkeit sein Buch mehrmals in unterschiedlichen Formaten herunterzuladen [ORe12b]

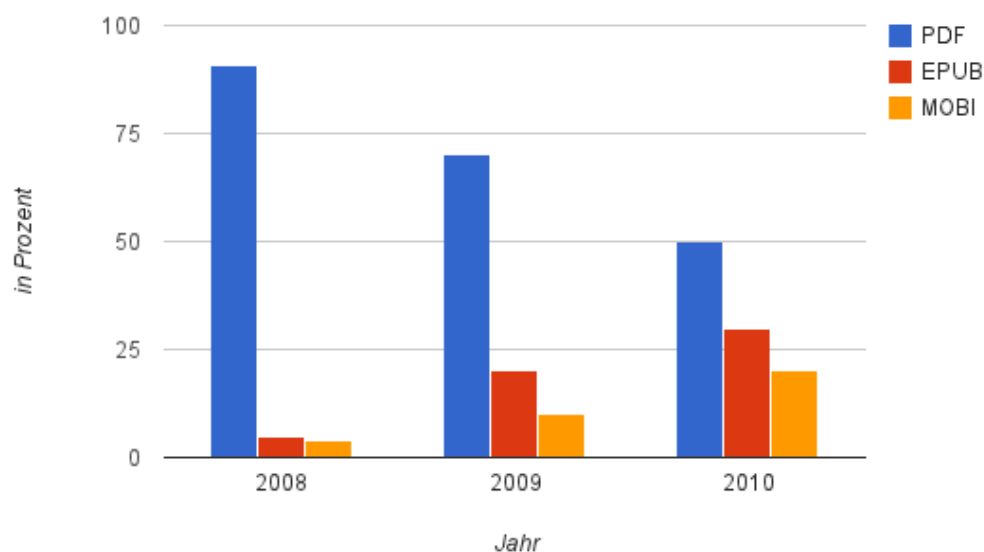


Abbildung 3.8: Downloadstatistik nach Formaten bei O'Reilly © O'REILLY

3.3 DRM (Digital Rights Management)

Die digitale Rechteverwaltung (DRM) ist eine Möglichkeit, die Verbreitung und Nutzung digitaler Inhalte zu kontrollieren. Die Kontrolle erfolgt meist durch eindeutige Kennzeichnung oder Verschlüsselung des Inhalts. [Gab12]

Eine größere Bekanntheit erlangte DRM mit dem Onlineverkauf von Musik. Um den illegalen Musik - Tauschbörsen ein legales Angebot entgegenzustellen, wurde im Jahr 2003 der iTunes Store von Apple gestartet. [Wik12h] Dieser Store ermöglichte auch den Kauf einzelner Lieder, welche aber mittels DRM geschützt waren. Diese Vorgehensweise wurde stark kritisiert, da die Nutzung der Lieder auch von den Käufern selbst nur sehr eingeschränkt und oft nur unter Problemen genutzt werden konnte.[Heio3] Aufgrund der Kritik wurde Anfang 2007 auch DRM freie Musik angeboten. Mittlerweile ist das gesamte Angebot im Store DRM frei. [Wik12g] . Im iTunes Store wird lediglich weiches DRM verwendet siehe Abschnitt 3.3.2 auf Seite 62.

Die Verwendung von DRM lässt sich in weiterer Folge in zwei Gruppen unterteilen. Einerseits gibt es eine Bezeichnung für »hartes DRM« und andererseits für »weiches DRM«. Beide verwenden unterschiedliche Ansätze, um eine Kontrollmöglichkeit zu gewährleisten.

3.3.1 Hartes DRM

Unter einem harten DRM versteht man Techniken und Funktionalität, die die Nutzung des digitalen Inhalts einschränken. Um diese Inhalte nutzen zu können, ist eine gewisse Abhängigkeit von externen Ressourcen wie z.B. Lizenzserver gegeben. Mittels hartem DRM lassen sich folgende Bereiche der Nutzung kontrollieren. [Lko8]

Nutzungsbeschränkung	Kopierfunktion Abspielfunktion Druckfunktion
Zugangsbeschränkung	Personen- oder Kontobezogener Zugang Endgerätebezogener Zugang Kombination aus beiden
Abrechnungssystem	Dauerhafte Nutzung nach einmaliger Abrechnung Temporäre Nutzung während eines Abonnements

Tabelle 3.1: Kontrollmöglichkeiten einzelner Bereiche mittels hartem DRM

Mit diesen Möglichkeiten könnten folgende Szenarien realisiert werden. [LKo8]

- Zeitliche Limitierung der Ressource (lässt sich nur 12 Monate ab Kauf öffnen)
- Beschränkung der Aufrufe (lässt sich nur 50x öffnen)
- Beschränkung der Ausdrücke
- Online / Offline Beschränkung (Aufruf nur mit Internetverbindung möglich)
- Beschränkung auf Hardware (Dokument ist nur auf einer Hardware (eReader) zu öffnen)

3.3.1.1 Grundlegende DRM Architektur für hartes DRM

Die Architektur eines DRM Systems besteht im Grunde aus drei Bereichen. Diese unterteilen sich wie folgt: [Sha12]

Content Server / Verleger / Herausgeber

Dieser Bereich ist in vier weitere Teile untergliedert. Er besteht aus einem Content-Repository, welches eine Meta Daten Management Umgebung bietet mit der der Gesamtbestand des Angebots eindeutig zuordenbar ist. Ein weiterer Teil sind Meta Daten über die digitale Ressource selbst wie z.B. Produktinformationen und Rechteinformation. Der dritte Teil ist ein sogenannter Packer, der den Inhalt mit den entsprechenden Mechanismen für den Endbenutzer verschlüsselt. Der letzte Teil ist eine Funktionalität zur Zustellung der Ressource an den Endbenutzer.

Lizenz Server / DRM Station

Der Lizenzserver bietet ein Schlüssel-Repository. Weiters bietet er eine Benutzerdatenbank, in der die Referenz-Identität zu Inhalt gespeichert wird. Um die Eindeutigkeit der Lizenzen zu gewährleisten, wird ein Lizenzgenerator betrieben. Dieser erstellt einerseits die Verbindung zwischen verschlüsseltem Inhalt und Identität her und andererseits die Verbindung zur Hardwareidentität des Endgerätes.

Client / User

Der Client befindet sich auf dem Endgerät des Nutzers und hat die Aufgabe den verschlüsselten Inhalt darzustellen. Weiters kommuniziert er mit dem Lizenzserver um die entsprechenden Rechte oder Berechtigungen zu setzen.

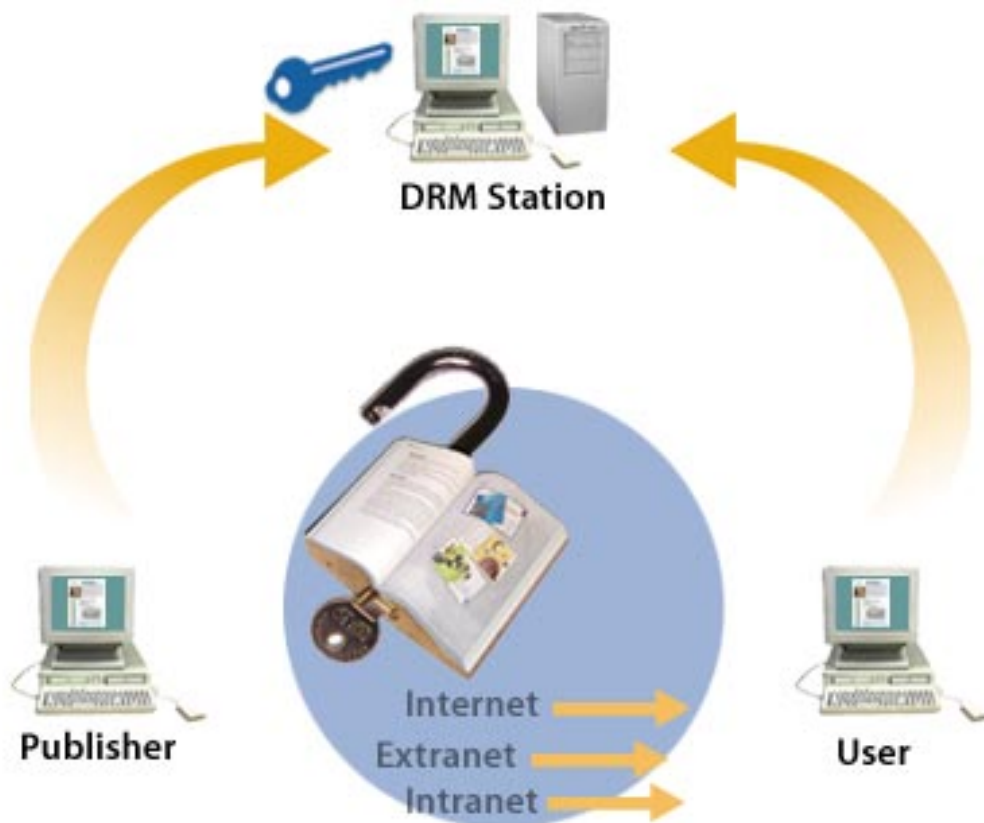


Abbildung 3.9: Grundlegende DRM Architektur [ORe12a]

3.3.1.2 Vorteile von hartem DRM

aus der Sicht der Hersteller und Verleger:[Pan10]

- Weitreichende Zugangs- und Nutzungskontrolle des angebotenen Inhaltes. Diese Kontrollmöglichkeiten sind auch noch direkt am Endgerät des Nutzers aktiv.
- Gewährleistung von Authentizität und Integrität

aus der Sicht der Nutzer und Konsumenten:

- Für den Nutzer und Konsumenten ergeben sich durch den Einsatz von hartem DRM keinerlei Vorteile

3.3.1.3 Nachteile von hartem DRM

aus der Sicht der Hersteller und Verleger:[Wik12c]

- Die Entwicklung und Bereitstellung eines DRM Systems ist sehr aufwendig und kostenintensiv
- Unzufriedenheit der Konsumenten und die Gefahr aufgrund der Restriktionen Kunden zu verlieren oder schlecht bewertet zu werden. Solche negativen Bewertungen müssen oft Hersteller von Computerspielen hinnehmen. So wurde im Jahr 2008 das Spiel Spore des Herstellers Electronic Arts aufgrund des übertrieben strengen DRM Systems sehr schlecht auf Amazon bewertet. Amazon entschied sich darauf hin über 2000 negative Bewertungen zu entfernen. [Walo8] Doch auch nach diesen Zensurschritten ist der derzeitige Stand so, dass von knapp 3000 Rezensionen 2600 Nutzer das Spiel aufgrund von DRM als sehr schlecht bewerten. [Ama12a]
- Verteuerung des Angebotes durch Weitergabe der Entwicklungs- und Instandhaltungskosten an den Konsumenten
- Kein Schutz gegen eine analoge Kopie. Der Schutz kann umgangen werden indem bei z.B. Musik die »analoge Lücke« ausgenutzt wird. Dabei wird der Ton mittels Mikrophon oder anderer Software aufgenommen und neu abgespeichert.
- Keine vollständige Sicherheit. Es gibt genug Programme und Techniken DRM zu umgehen. Ist der Schutz einmal ausgehebelt, müsste wieder an Verbesserungen gearbeitet werden, was wiederum mit Kosten verbunden ist.

aus der Sicht der Nutzer und Konsumenten:[Wik12c]

- Einschränkungen in der Nutzung. Das kann soweit gehen, dass ein mit dem Desktop Computer im Internet gekauftes Buch nur auf diesem lesbar ist. Es ist dann nicht möglich, das Buch auf einem Laptop oder eReader zu kopieren und es dort zu verwenden.
- Inkompatibilität verschiedener Systeme. Da jeder Hersteller oft sein eigenes System verwendet, muss auch das Endgerät mit diesen Systemen umgehen können. Dies ist leider sehr oft nicht der Fall.

- Fremdkontrolle der Anbieter über bereits gekaufte Inhalte. Der Hersteller Amazon nutzte diese Fähigkeit, um im Nachhinein einen bereits gekauften Buchtitel nachträglich direkt am Endgerät des Nutzers zu löschen. Ironischerweise waren das die Titel »1984« und »Animal Farm« von »George Orwell«, welcher in den Büchern die Zensur und Fremdkontrolle thematisiert. [Sto09]
- Aufwändigere Datensicherung. Es muss zusätzlich zur digitalen Ressource auch die Lizenzinformation abgespeichert werden.
- Durch die fehlende Kompatibilität von DRM Implementierungen wird der Nutzer an ein DRM System gebunden. Der freie Markt steht dem Nutzer dabei nicht mehr offen.

3.3.2 Weiches DRM

Bei weichem DRM passiert keinerlei Einschränkung mittels technischen Hilfsmitteln. Diese Technik basiert auf Wasserzeichen. Diese Wasserzeichen können sichtbar oder unsichtbar in einer digitalen Ressource abgelegt werden. Ziel dabei ist es, einen psychologischen Effekt zu nutzen. Die digitale Ressource wird beim Kauf personalisiert. Diese Personalisierung kann z.B. durch das Anzeigen von Kontoinformationen des Käufers am Produkt selbst erfolgen. [Sch10]

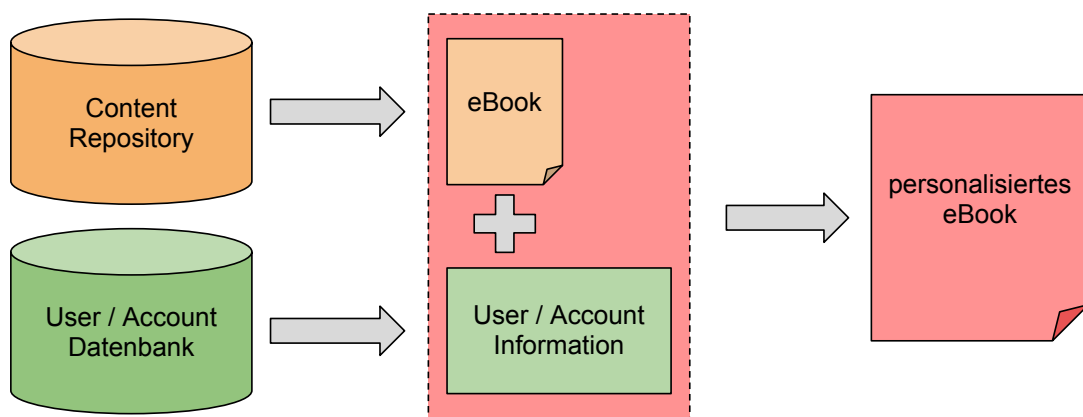


Abbildung 3.10: Grundlegender Personalisierungsvorgang bei weichem DRM

Diese Technik wird aktuell auch beim »DRM freien« iTunes Store von Apple verwendet. Hier wird jeder heruntergeladenen Datei eine personalisierte Kontoinformation beigefügt.[Wik12g]

ßen. Rechteinhaber haben dadurch die Möglichkeit im Nachhinein zu überprüfen, ob Teile der verkauften Ressourcen nicht wie vorgesehen verwendet werden. Wie mit dieser Information aber weiter umgegangen wird liegt dabei im Ermessen des Auftraggebers. Eine weitere positivere Eigenschaft von Wasserzeichen ist, dass diese für Studien verwendet werden können. Wenn ein Verleger in einem begrenzten Zeitraum kostenlose eBooks verteilt, kann er deren weitere Verbreitung sehr genau nachvollziehen. Die daraus gewonnene Erkenntnis kann in weiterer Folge für Werbezwecke genutzt werden.[SWD09]

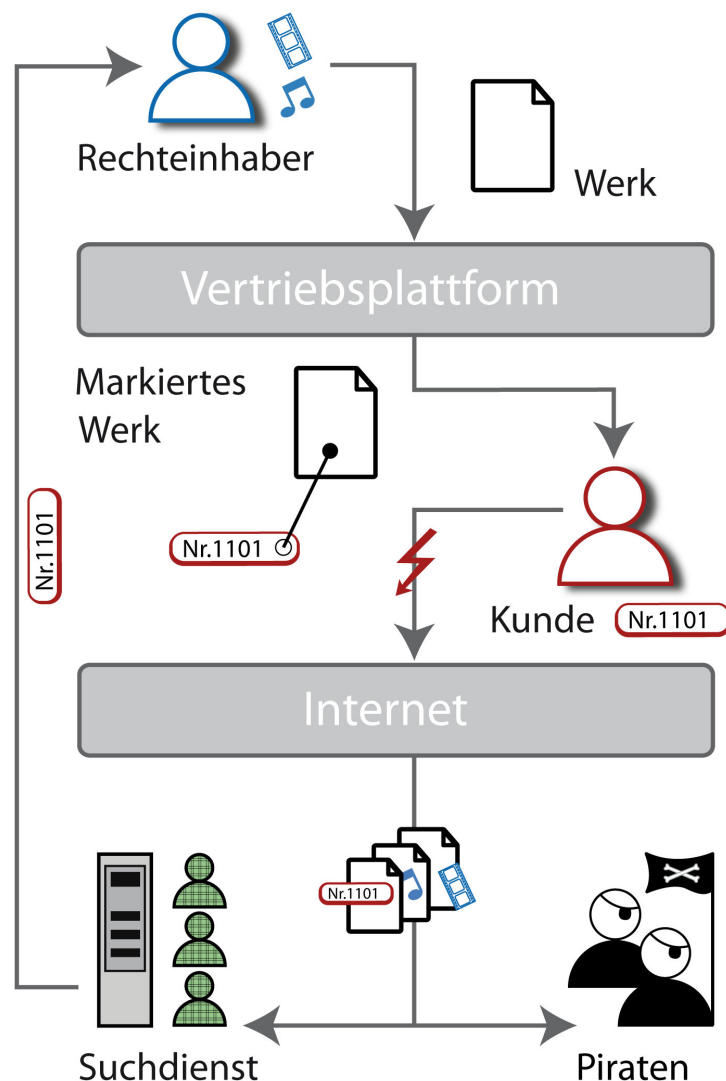


Abbildung 3.12: Suche nach illegalen Werken mit Hilfe von Wasserzeichen [SWD09]

3.3.2.1 Vorteile von weichem DRM

aus der Sicht der Hersteller und Verleger:

- Bessere Kundenakzeptanz, da keinerlei technische Kontrollmöglichkeiten und Einschränkungen bei der Nutzung entstehen.
- Nachträgliche Möglichkeit Missbrauch zu entdecken und verantwortliche Nutzer oder Konten zu sperren.

aus der Sicht der Nutzer und Konsumenten:

- Für den Nutzer und Konsumenten ergeben sich durch den Einsatz von weichem DRM keinerlei Vorteile.

3.3.2.2 Nachteile von weichem DRM

aus der Sicht der Hersteller und Verleger:[Sch10]

- Keine direkte Zugangs,- und Nutzungskontrolle.
- Missbräuchliche Verwendung kann erst im Nachhinein festgestellt werden.
- Rechtliche Aspekte sind nicht klar definiert. Folgendes Beispiel: Ein Nutzer verliert seinen USB-Stick mit den darauf gespeicherten personalisierten Medien. Diese Medien werden von einem unehrlichen Finder illegal hochgeladen. Wer trägt hier die Verantwortung? Kann hier der Nutzer, der die Medien gekauft hat, zur Verantwortung gezogen werden? Wohl kaum. Daraus ergibt sich eine schwierige Beweisführung.
- Nachforschen der Medien und Überprüfung ist aufwendig und mit Kosten verbunden. Für ständige und detaillierte Nachforschungen sind oft externe Dienstleister notwendig.
- Fehlende Kundenakzeptanz bei Nutzern mit Datenschutzbedenken.

aus der Sicht der Nutzer und Konsumenten:[Sch10]

- Probleme mit dem Datenschutz. Wenn Kontoinformation wie z.B. Emailadressen im Klartext abgespeichert werden, könnten diese missbräuchlich verwendet werden. Die Gewissheit, dass jedes gekaufte Medium eindeutig zuordenbar ist, könnte einige Nutzer vom Kauf abhalten.

3.3.3 Aktueller DRM Einsatz

Anfang des Jahres 2011 wurde deutschlandweit eine Umfrage unter Experten des Verlagswesens durchgeführt. Ziel dieser Befragung war herauszufinden, welche technischen Schutzmaßnahmen die Verlage nutzen, um eine unautorisierte Weiterverbreitung von eBooks zu unterbinden.

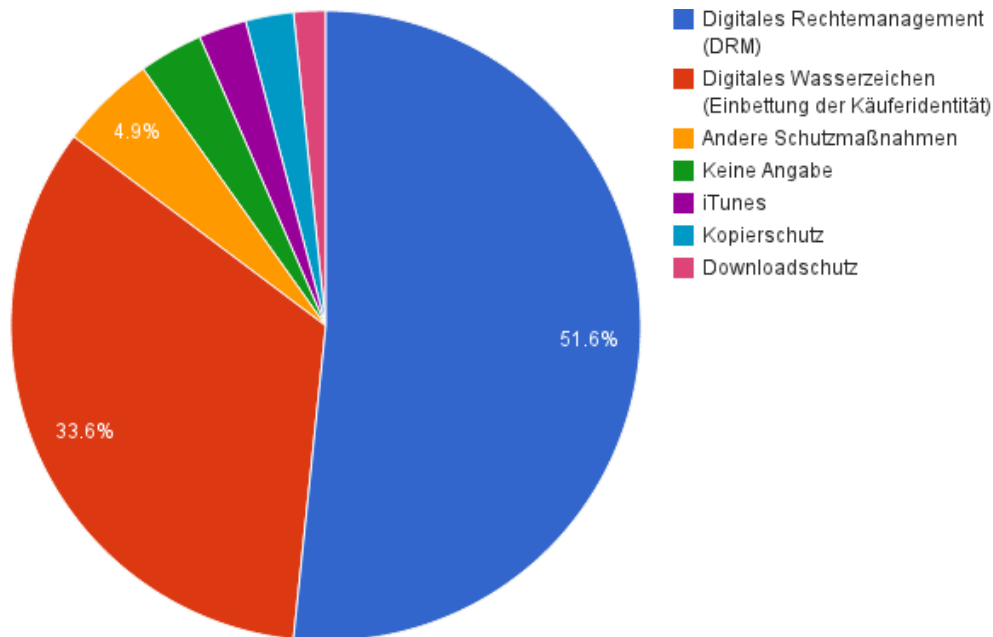


Abbildung 3.13: Schutzmaßnahmen von Verlagen gegen Piraterie bei E-Books © STATISTA

Im Jahr 2010 wurden Konsumenten zu dem Thema DRM befragt. Man wollte herausfinden, welche Punkte den größten Widerstand hervorrufen. Dazu konnten die Nutzer aus mehreren Fragen wählen.

Ich kann E-Books nicht an Freunde/Bekannte ausleihen	61%
Ich kann E-Books nicht auf mehreren Geräten lesen	60%
Die E-Books sind deshalb eventuell nicht auf allen Geräten lesbar	57%
Bestimmte Rechte wie Ausdrucken etc. können eingeschränkt sein	51%
Ich kann E-Books nicht weiterverkaufen	32%
Es gibt bessere Möglichkeiten, die unkontrollierte Weitergabe von E-Books zu verhindern	21%
keine dieser Möglichkeiten	12%

Tabelle 3.2: Kundenmeinung zur DRM bei E-Books und E-Readern © STATISTA

3.4 Metadaten

Metadaten beschreiben andere Daten. Mit Hilfe von Metadaten lassen sich Eigenschaften und Informationen von anderen Daten beschreiben. Metadaten sind mindestens gleich wichtig und wertvoll wie die beschriebenen Daten selbst. Vor allem in Bereichen wo sich größere Datenmengen sammeln sind Metadaten essentiell, um hier einen Überblick wahren zu können. Der Begriff Metadaten ist zwar relativ neu und wird eher dem digitalen Zeitalter zugeschrieben, doch wurden diese schon früher in Form von z.B. Bibliothekskatalogen angelegt und verwendet. [Wik12i]. Eine Sammlung von Metadaten in den Bereichen Bibliothek (orange) und Archiv (violett) zeigt untere Grafik. Die Intensität der Farbe zeigt hier, wie stark der Metadatenstandard an das Einsatzgebiet z.B. Bibliothek gebunden ist.

AACR2, AGLS, CQL, DDC, FRAD, FRBR, FRASAD, GILS, ISBD, LCC, LCSH, MADS, MARC, MARC Relator Codes, MARCXML, MESH, METS, MIX, MODS, OAI-PMH, OAI, OpenURL, PREMIS, RDA, Sears List of Subject Headings, SRU, SWAP, TEI, TextMD, TGM I, TGM II, VRA Core, XML, XML Schema, XOBIS, XPath, XSLT, Z39.50	AGLS, APPM, DACS, EAC-CPF, EAD, GILS, ISAAR(CPF), ISAD(G), RAD
AES Core Audio, AES Process History, CanCore, CCO, DC, DCAM, DTD, FGDC/CSDGM, GEM, IEEE/LOM, MEI, METS Rights, OAI-ORE, PB Core, QDC, RDF, SGML, TGN, XQuery	DTD, LCC, LCSH, MARC, MARCXML, METS, MIX, MODS, OAI-PMH, OAI, PB Core, PREMIS, SGML, SRU, TGM I, TGM II, TGN, XML, XML Schema, XPath, XQuery, XSLT
APPM, Atom, CDWA, CDWA Lite, CIDOC/CRM, DACS, DwC, EAC-CPF, EAD, EML, FOAF, indecs, ISAAR(CPF), ISO 19115, Linked Data, MPEG-21 DIDL, ONIX, RELAX NG, RSS, SKOS, Topic Maps, ULAN	AES Core Audio, Atom, CIDOC/CRM, DC, DCAM, FGDC/CSDGM, FOAF, FRAD, FRBR, FRASAD, ISO 19115, Linked Data, OAI-ORE, QDC, RDF, RELAX NG, RSS, SKOS, TEI, Topic Maps, VRA Core, XOBIS
AAT, ADL, DIF, ID3, ISAD(G), KML, MPEG-7, MusicXML, MXF, ODRL, RAD, SMIL, VSO Data Model, XMP, XRML	AACR2, AAT, ADL, CanCore, CDWA, CDWA Lite, DDC, DwC, GEM, IEEE/LOM, indecs, ISBD, KML, MADS, MESH, METS Rights, MPEG-7, ODRL, RDA, SMIL, TextMD, ULAN, VSO Data Model, XMP, XrML, Z39.50

Abbildung 3.14: Verwendete Metadatenstandards in den Bereichen Bibliothek (orange) und Archiv (violett) [RB10]

Die Erstellung von Metadaten kann unterschiedliche Bereiche betreffen. Darum werden drei Grundtypen von Metadaten unterschieden. [NISO4] Diese Grundtypen sind nicht immer eindeutig voneinander abgrenzbar, sodass auch Überschneidungen auftreten können.

3.4.1 Deskriptive Metadaten

Mit deskriptiven Metadaten wird die digitale Ressource selbst beschrieben. Durch diese Beschreibung erhält man im besten Fall Antworten auf folgende Fragen.

- Wer hat diese Ressource erstellt? (Name des Autors)
- Was ist der Inhalt der Ressource? (Titel eines Buches)
- Wann wurde die Ressource erstellt? (Erscheinungsjahr)
- Wo (von Wem) wurde die Ressource publiziert? (Verlag)

Die Beschreibung kann unter anderem mit definierten Schlagwörtern unterstützt werden. Durch solche Kataloge an Wörtern lassen sich auch Beziehungen zu anderen Objekten aufbauen und darstellen. Mit Hilfe dieser Daten ist auch eine eindeutige Identifikation möglich. [Arm99]

3.4.2 Strukturelle Metadaten

Strukturelle Metadaten beschreiben Eigenschaften, wie die digitale Ressource dem Nutzer zugänglich gemacht werden kann. Es betrifft hierbei den Aufbau eines z.B. Dokumentes wie kann man sich durch dieses Dokument navigieren, welche Nutzungsmöglichkeiten ergeben sich aus dem Aufbau. Eine Angabe kann hier z.B. die Anzahl der Seiten oder Anzahl der Bilder sein sein. Weiters werden Informationen über die angebotenen oder verfügbaren Formate gespeichert. [Hur+99]

3.4.3 Administrative Metadaten

Administrative Metadaten dienen vorwiegend der Verwaltung. Diese können z.B. zur Verrechnung, Kategorisierung oder Archivierung verwendet und angelegt werden. Gerade wenn ältere Ressourcen aufwendig digitalisiert werden, fallen administrative Metadaten wie z.B. Scandatum, Scanauflösung, verwendete Hardware, Dateiname, usw. an. [Hur+99]

3.4.4 Metadatenstandards

Es gibt mittlerweile eine Menge an Metadatenstandards. Jeder Standard deckt einen gewissen Bereich von Anforderungen ab. Die weitere Auswahl von Metadaten betreffen hauptsächlich die Bereiche Bibliothek und Archiv. Untere Grafik bietet einen Überblick über die in den letzten Bereichen verwendeten Standards.

3.4.4.1 DC (Dublin Core)

Dublin Core wurde im Jahr 1995 auf einer Konferenz in Ohio/Dublin entwickelt. Ziel war es, ein System oder Schema zu entwickeln, um einen Großteil der vorhandenen Ressourcen einfach und schnell beschreiben zu können. Das Dublin Core Metadaten Element Set besteht aus 15 Eigenschaften. Die Elemente sind alle optional und können auch mehrfach verwendet werden. Die Weiterentwicklung und Wartung von Dublin Core übernimmt hierbei die Dublin Core Metadata Initiative (DCMI). Mit Hilfe dieser 15 Kernelemente lassen sich Informationen über den Inhalt, den technischen Eigenschaften, Personen, Rechte und weiteren Referenzen pflegen.[dub10].

Die folgende Tabelle zeigt die 15 Kernelemente von Dublin Core.

Title	Titel der beschriebenen Ressource
Creator	Hauptverantwortlicher für den Inhalt
Subject	Schlüsselwort für den Inhalt
Description	Beschreibung des Inhalts
Publisher	Verantwortlicher für die Veröffentlichung der Ressource
Contributor	Beteiligter am Entstehen der Ressource
Date	Entstehungsdatum
Type	Art der Ressource
Format	Format der Ressource
Identier	Einzigartiger Bezeichner zur Identifizierung
Source	Verweis auf andere Ressource die der beschriebenen zugrunde liegt
Language	Sprache des Inhalts
Relation	Beziehung zu einer anderen Ressource
Coverage	zeitliche und räumliche Eigenschaften der Ressource
Rights	Informationen zum Rechtemanagement

Tabelle 3.3: Kernelemente von Dublin Core [dub10]

Beispielhafte Dublin Core Codierung im RDF (Resource Description Framework).

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">

  <rdf:Description rdf:about="http://wikipedia.org/wiki/">
    <dc:creator> Wikimedia Foundation Inc.</dc:creator>
    <dc:title>Wikipedia, the free encyclopedia</dc:title>
    <dc:description>The name Wikipedia is a portmanteau of the words wiki (a technology for creating collaborative websites, from the Hawaiian word wiki, meaning quick) and encyclopedia.
    </dc:description>
    <dc:date>2012-05-29</dc:date>
  </rdf:Description>
</rdf:RDF>
```

Quellcode 3.6: Beispielhafte Codierung mit Dublin Core in RDF/XML [dub10]

3.4.4.2 METS (Metadata Encoding & Transmission Standard)

METS ist ein Standard zur Beschreibung von deskriptiven, strukturierten und administrativen Metadaten. METS verwendet XML für die Syntax und das Vokabular. METS ist ein XML Schema, was entwickelt wurde, um XML Dokumente zu erstellen, um damit den hierarchischen Aufbau eines digitalen Objekts zu beschreiben. METS ist ein offener Standard, erweiterbar und modular. Die grundlegende Struktur von METS ist in sieben Bereiche unterteilt. [Con12c]

METS Header	metsHdr	Metadaten über das METS Dokument
Descriptive Metadata	dmdSec	Die Erschließungsdaten des Objekts
Administrative Metadata	amdSec	Enthält Verwaltungsinformation wie Ursprung oder Urheberrechte
File Section	fileSec	Enthält Information über alle zusätzlichen Objekte
Structural Map	structMap	Beschreibt die interne Struktur des Objekts
Structural Links	structLink	Beinhaltet Verknüpfungen wie z.B. Archivierung von Webseiten
Behavioral	behaviorSec	Dieser dient dazu, um ausführbare Elemente wie z.B. Webservices mit dem METS-Paket zu verbinden

Tabelle 3.4: Grundstruktur von METS [Con12c]

Innerhalb dieser METS Struktur können unterschiedliche Metadatenstandards verwendet werden. Unten angeführter Code zeigt, wie in einem METS Dokument der Bereich Descriptive Metadata mittels Dublin Core beschrieben wird.

```
<dmdSec ID="dmd002">
  <mdWrap MIMETYPE="text/xml" MDTYPE="DC" LABEL="Dublin Core Metadata">
    <dc:title>Alice's Adventures in Wonderland</dc:title>
    <dc:creator>Lewis Carroll</dc:creator>
    <dc:date>between 1872 and 1890</dc:date>
    <dc:publisher>McCloughlin Brothers</dc:publisher>
    <dc:type>text</dc:type>
  </mdWrap>
</dmdSec>
```

Quellcode 3.7: METS Dokument mit Dublin Core Tags [Con12c]

3.4.4.3 MARC (Machine Readable Cataloging)

MARC wurde im Jahr 1960 von der Library of Congress entwickelt und war damit eines der ersten elektronischen Metadatenformate. Ziel der Entwicklung war es, einen leichteren elektronischen Austausch von Metainformation zwischen Bibliotheken zu ermöglichen. Mit MARC kann sehr viel sehr genau beschrieben werden. Durch die Verwendung von umfangreichen Zahlencodes kann die Verwendung von MARC sehr komplex werden. [Con12a]

001	Control Number
003	Control Number Identifier
005	Date and Time of Latest Transaction
041	Language Code
042	Authentication Code
044	Country of Publishing/Producing Entity Code
045	Time Period of Content
080	Universal Decimal Classification Number
082	Dewey Decimal Classification Number
256	Computer File Characteristics
257	Country of Producing Entity
504	Bibliography, Etc. Note
505	Formatted Contents Note
506	Restrictions on Access Note
856	Electronic Location and Access
880	Alternate Graphic Representation

Tabelle 3.5: Auszüge von Zahlencodes in MARC) [Con12b]

Untere Grafik zeigt einen Beispieldatensatz wie er im MARC codiert ist.

```
00731nam a22002295a
4500001000900000005001700009008004100026906004500067925004400112955002000156
0100017001760200030001930200025002230400013002481000020002612450043002812500
0120032426000520033626300090038830000110039796300930040815015399200709201630
36.0070920s2007 wyu 000 0 eng a0bibccorignewd2eepcnf20gy-gencatlg0
aacquireb2 shelf copiespolicy default apc21 2007-09-20 a 2007937909 a9781605120645
(hardcover) a9781605121642 (pbk.) aDLCcDLC1 aOrwell, George.10aNineteen eighty four
/cGeorge Orwell. a1st ed. aJackson Hole, WY :bArcheion Press, LLC,c2007. a0709 ap. cm.
aEditor; phone: (877) 745-7317; email: lccn@archeionpress.com; bc: lccn@archeionpress.com
```

Abbildung 3.15: MARC Datensatz codiert (Library of Congress Online Catalog)

```
LC Control No.: 2007937909
LCN Permalink: http://lccn.loc.gov/2007937909
000 00731nam a22002295a 450
001 15015399
005 20070920163036.0
008 070920s2007 wyu 000 0 eng
906 __ |a 0 |b ibc |c orignew |d 2 |e epcn |f 20 |g y-gencatlg
925 0_ |a acquire |b 2 shelf copies |x policy default
955 __ |a pc21 2007-09-20
010 __ |a 2007937909
020 __ |a 9781605120645 (hardcover)
020 __ |a 9781605121642 (pbk.)
040 __ |a DLC |c DLC
100 1_ |a Orwell, George.
245 10 |a Nineteen eighty four / |c George Orwell.
250 __ |a 1st ed.
260 __ |a Jackson Hole, WY : |b Archeion Press, LLC, |c 2007.
263 __ |a 0709
300 __ |a p. cm.
963 __ |a Editor; phone: (877) 745-7317; email: lccn@archeionpress.com;
```

Abbildung 3.16: MARC Datensatz aufbereitet (Library of Congress Online Catalog)

Darauf aufbauend wurde im Jahr 2002 MARC-XML entwickelt. MARC-XML erlaubt die Darstellung von MARC Datensätzen in XML. Dadurch können Bibliotheken ihre Datensätze über ein Webservice zur Verfügung stellen. [Con12b]

3.4.4.4 MODS (Metadata Object Description Schema)

Die Entwicklung von MODS begann im Jahr 2002 und befindet sich von der Komplexität zwischen MARC und Dublin Core. Es ist ein XML Schema, welches von der Library of Congress gewartet und weiter entwickelt wird. Für jede Eigenschaft gibt es ein XML Tag. Diese Tags sind teilweise zu MARC kompatibel. MODS hat mehr Elemente wie Dublin Core aber weniger Elemente als MARC zur Verfügung. Trotzdem ist eine teilweise Konvertierung mittels XSLT möglich. Diese Scripts werden von der Library of Congress zur Verfügung gestellt. MODS bietet zur Beschreibung 19 Top Level Elemente. [Con12d]

titleInfo	note
name	subject
typeOfResource	classification
genre	relatedItem
originInfo	identifier
language	location
physicalDescription	accessCondition
abstract	extension
tableOfContents	recordInfo
targetAudience	

Tabelle 3.6: MODS Top Level Elemente [Con12d]

```
<?xml version="1.0" encoding="UTF-8"?>
<modsCollection xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/
  XMLSchema-instance" xmlns="http://www.loc.gov/mods/v3" xsi:schemaLocation="http://www.
  loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-3.xsd">
<mods version="3.3">
<titleInfo>
  <title>1984</title>
</titleInfo>
<name type="personal">
  <namePart>George Orwell</namePart>
</name>
<typeOfResource>text</typeOfResource>
<originInfo>
  <publisher>New American Library</publisher>
</originInfo>
<language>
  <languageTerm type="text">English</languageTerm>
</language>
<physicalDescription>
  <digitalOrigin>reformatted digital</digitalOrigin>
</physicalDescription>
</mods>
</modsCollection>
```

Quellcode 3.8: MODS - Codierte Metadaten [Con12d]

3.4.4.5 FRBR (Functional Requirements for Bibliographic Records)

FRBR ist weniger ein Standard für Metadaten sondern ein abstraktes relationales Modell. FRBR basiert auf einem Entity-Relationship Modell. Es werden dabei Konzepte und deren Beziehung zueinander erstellt. Es wird versucht, dadurch Definitionen zu erstellen, auf denen sich andere Regelwerke abbilden lassen. Grund der Entwicklung von FRBR war die Navigation in umfangreichen Bibliothekskatalogen zu vereinfachen um dadurch die Bedürfnisse der Nutzer besser erfüllen zu können. Der grundlegende Aufbau von FRBR besteht aus drei Gruppen in denen jeweils Entitäten und Relationen zusammengefasst sind. [Blo+05]

Gruppe 1	Work Expression Manifestation
Gruppe 2	Item Person Corporate Body
Gruppe 3	Concept Object Event Place

Tabelle 3.7: FRBR Entitäten Gruppen [Mad+09]

Mit den Entitäten in Gruppe 1 werden die geistigen und künstlerischen Aspekte des Werkes beschrieben.

- Die Entität »Work« ist dabei eine abstrakte Beschreibung z.B. ein Roman (1984).
- Die Entität »Expression« beschreibt die Eigenschaften, mit der das Werk tatsächlich realisiert wurde z.B. eine deutsche Übersetzung oder mündliche Überlieferung (Deutsche Übersetzung von Kurt Wagenseil).
- Mit der Entität »Manifestation« beschreibt man die Erscheinungsform des Werkes z.B. als Buch, als CD-ROM, Zeit der Veröffentlichung (36. Auflage 2002).
- Die Entität »Item« beschreibt dann ein konkretes Exemplar eines Buches (Ein Exemplar zur Ausleihe in der Stadtbibliothek Graz mit der Buchungsnummer 0202BU19027).[Blo+05]

Untere Grafik zeigt das oben genannte Beispiel und deren Verbindungen zwischen den Entitäten.

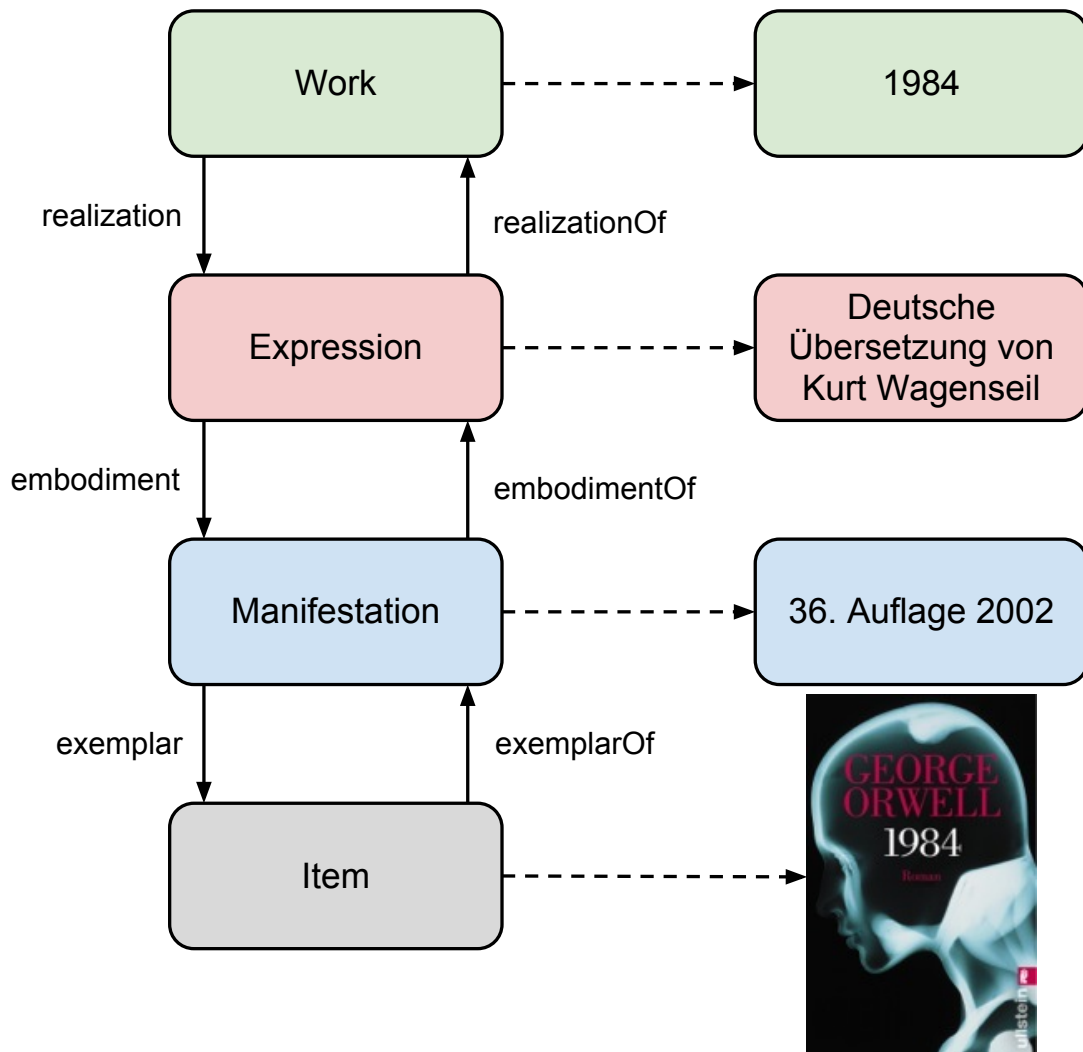


Abbildung 3.17: FBFR Entitäten der Gruppe 1 mit Beispiel [Mad+09]

Die Entitäten in Gruppe 2 repräsentieren diejenigen, die für die geistigen und künstlerischen Aspekte des Werkes verantwortlich sind. Darüber hinaus wird auch Information über die Verwaltung, Weitergabe und die physische Erstellung abgelegt. Dabei kommen die Entitäten »Person« und »Corporate Body« zum Einsatz. [Mad+09]

Die Entitäten in Gruppe 3 bieten zusätzliche Information über das Werk als

Gegenstand an. Dazu werden die Entitäten »Concept«, »Object«, »Event« und »Place« verwendet. [Mad+09]

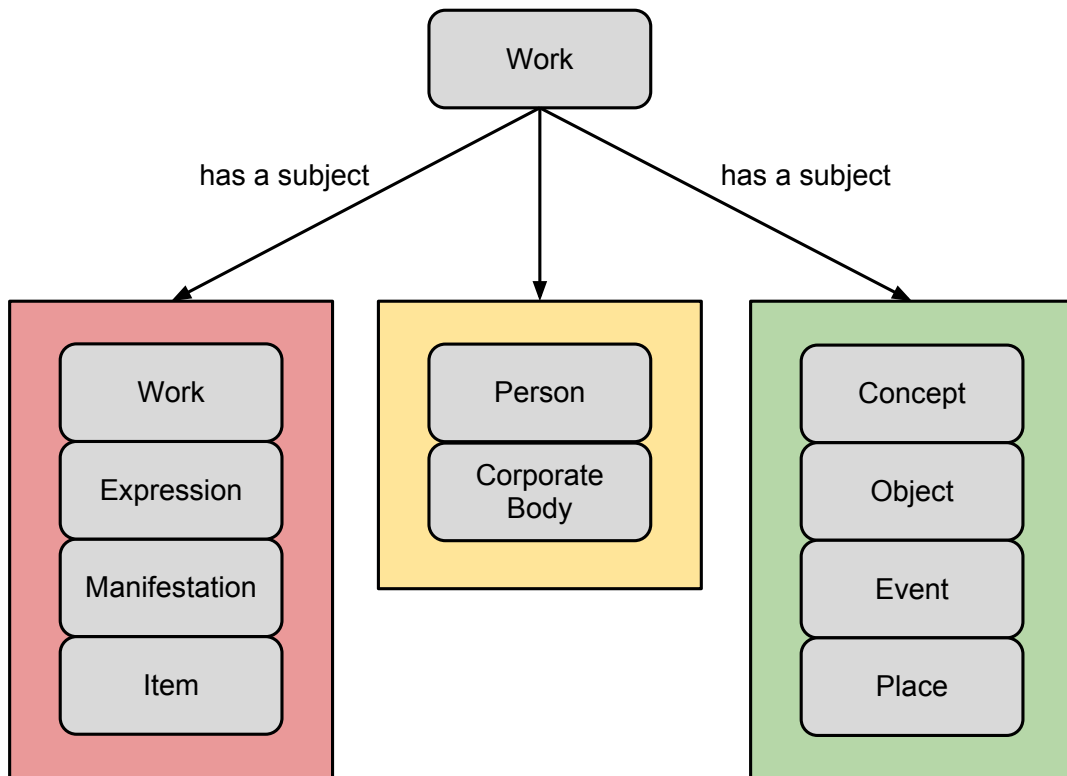


Abbildung 3.18: Alle FRBR Gruppen und die Gegenstands (subject) Beziehungen [Mad+09]

3.5 Austauschprotokolle

3.5.1 Z39.50

Das Protokoll Z39.50 wurde im Jahr von der NISO (National Information Standards Organization) entwickelt, um einen Standard für Abfragen im Bibliothekswesen zu erstellen. Es ermöglicht eine systemübergreifende Suche und den Austausch bibliographischer Daten. Dabei ist es irrelevant in welchem System und in welcher Struktur die Daten in den einzelnen Bibliotheken abgelegt sind. Seit 1990 wird das Protokoll von der Library of Congress weiter betreut. Z39.50 ist

seit 1998 Standard ISO23950. Z39.50 verwendet zu Abfrage ein Set von definierten Wörtern und eine Syntax, um boolesche Logik einzubinden. [Con11]

3.5.1.1 SRU/CQL

Mittlerweile gibt es einige Weiterentwicklungen, die auf Z39.50 aufbauen. Diese sind z.B. SRU/CQL. SRU (Search Retrieval via Url) ermöglicht Abfragen über HTTP GET und SOAP (Simple Object Access Protocol). Die Anfragen werden dabei mit der Sprache CQL (Common Query Language) ausgedrückt und das Ergebnis als XML zurückgeliefert. [Con12e]

3.5.2 OAI-PMH

Das OAI-PMH (OAI Protocol for Metadata Harvesting) ist ein Protokoll zum Verarbeiten von MetaDaten. Es wurde von der OAI (Open Archives Initiative) bereit gestellt und entwickelt. Ziel war es, die bibliographischen Daten von Publikationen, die auf Servern abgelegt waren, besser nutzen zu können. OAI-PMH basiert auf den Technologien XML und REST (Representational State Transfer). [Lag+02] Das Protokoll dient primär zum Sammeln von Metadaten. Diese Metadaten werden von Data Providern zur Verfügung gestellt. Data Provider können mehrere Metadatenformate anbieten, müssen aber mindestens Dublin Core verwenden. Die von den Data Providern zur Verfügung gestellten Metadaten werden in weiterer Folge von Service Providern aufbereitet und dem Nutzer zur Verfügung gestellt. [OAI12].

OAI unterstützt folgende Befehle:

GetRecord	Liefert einen Metadatenatz mittels mitgegebenen Identifier und dem geforderten Metadatenpräfix
Identify	Liefert Informationen über den Data Provider
ListIdentifiers	Liefert eine Liste von Identifiers
ListMetadataFormats	Liefert eine Liste der Metadaten die der Data Provider unterstützt bzw. anbietet
ListRecords	Liefert eine Liste von Datensätzen zum Sammeln von Metadaten
ListSets	Liefert den Aufbau, die Struktur des Data Providers

Tabelle 3.8: OAI Befehlssatz [Lag+02]

Um nun festzustellen, welche Metadaten ein Data Provider unterstützt, ist folgender Aufruf notwendig. Zum Test dient in diesem Fall der Data Provider Università degli Studi di Perugia in Italien.

OAI Request - ListMetadataFormats:

```
http://www.rei.unipg.it/rei/oai?verb=ListMetadataFormats
```

Antwort codiert in XML:

```
<MetadataFormats>
  <MetadataFormat>
    <Prefix>oai_dc</Prefix>
    <Schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd</Schema>
    <Namespace>http://www.openarchives.org/OAI/2.0/oai_dc</Namespace>
  </MetadataFormat>
  <MetadataFormat>
    <Prefix>oai_marc</Prefix>
    <Schema>http://www.openarchives.org/OAI/1.1/oai_marc.xsd</Schema>
    <Namespace>http://www.openarchives.org/OAI/1.1/oai_marc</Namespace>
  </MetadataFormat>
  <MetadataFormat>
    <Prefix>marcxml</Prefix>
    <Schema>http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd</Schema>
    <Namespace>http://www.loc.gov/MARC21/slim</Namespace>
  </MetadataFormat>
  <MetadataFormat>
    <Prefix>nlm</Prefix>
    <Schema>http://dtd.nlm.nih.gov/publishing/2.3/xsd/journalpublishing.xsd</Schema>
    <Namespace>http://td.nlm.nih.gov/publishing/2.3</Namespace>
  </MetadataFormat>
  <MetadataFormat>
    <Prefix>rfc1807</Prefix>
    <Schema>http://www.openarchives.org/OAI/1.1/rfc1807.xsd</Schema>
    <Namespace>http://info.internet.isi.edu:80/in-notes/rfc/files/rfc1807.txt</Namespace>
  </MetadataFormat>
</MetadataFormats>
```

Quellcode 3.9: Angebotene Metadaten nach einem OAI Request

3.5.3 OPDS

OPDS (Open Publication Distribution System) ist ein Katalogformat für elektronische Medien. OPDS basiert auf dem Atom Syndication Format. [Soc05] Dieses Format ist einfach im Aufbau und garantiert eine hohe Geschwindigkeit. Die Spezifikation von OPDS beschreibt den feed-basierten Aufbau dieser Kataloge. Die Kataloge ermöglichen es, elektronische Medien zu verwalten, diese such- bzw. findbar zu machen und deren Verbreitung zu unterstützen. Da die Spezifikation von OPDS auf etablierte Standards wie Atom und HTTP aufbaut, ist auch eine breite Unterstützung gegeben. [OPD12a]

Bekannte Anbieter von OPDS Katalogen sind unter anderem: [OPD12b]

- Projekt Gutenberg
<http://m.gutenberg.org>
- Feedbooks
<http://www.feedbooks.com/catalog.atom>
- Internet Archive
<http://bookserver.archive.org/catalog/>
- O'Reilly Media
<http://opds.oreilly.com/opds/>
- u.v.m.

Unter Punkt 7, der Spezifikation von OPDS Catalog 1.1, werden grundsätzlich zwei Feeds unterschieden.

- Navigations-Feeds
- Aquisitions-Feeds

Navigations-Feeds dienen dazu, eine Struktur mit Hierarchien innerhalb eines Kataloges aufzubauen. Die Aquisitions-Feeds zeigen Möglichkeiten auf, wie das Buch erworben werden kann. [OPD12a]

Als Beispiel wird hier der OPDS Katalog von O'Reilly gezeigt.

```
<feed xmlns="http://www.w3.org/2005/Atom">
  <id>http://opds.oreilly.com/opds/</id>
  <link rel="self" href="/opds/" type="application/atom+xml;profile=opds-catalog" />
  <link rel="start" href="/opds/" type="application/atom+xml;profile=opds-catalog" />
  <title>O'Reilly Media Catalog</title>
  <updated>2010-01-10T10:01:10Z</updated>
  <author>
    <name>O'Reilly Media</name>
  </author>
  <entry>
    <title>New Releases</title>
    <link href="/opds/new" rel="http://opds-spec.org/sort/new" type="application/atom+xml;
      profile=opds-catalog" />
    <updated>2010-01-10T10:01:01Z</updated>
    <id>http://opds.oreilly.com/new</id>
  </entry>
  <entry>
    <title>Bestsellers</title>
    <link href="/opds/best" rel="http://opds-spec.org/sort/popular" type="application/atom+xml;
      profile=opds-catalog" />
    <updated>2010-01-10T10:01:01Z</updated>
    <id>http://opds.oreilly.com/best</id>
  </entry>
  <entry>
    <title>Alphabetical</title>
    <link href="/opds/alpha" rel="subsection" type="application/atom+xml;profile=opds-catalog" />
    <updated>2010-01-10T10:01:01Z</updated>
    <id>http://opds.oreilly.com/alpha</id>
  </entry>
</feed>
```

Quellcode 3.10: O'Reilly OPDS Katalog

Wie im Code ersichtlich, ist der Katalog von O'Reilly in drei weitere Aufrufe unterteilt:

- New Releases
- Bestsellers
- Alphabetical

Hinter diesen Aufrufen befindet sich dann die detaillierte Information über die Bücher, die der Kategorie zugeordnet sind. Um die Navigation und Bedienung besser darstellen zu können, wurde ein Reader auf einem Smartphone installiert. Dazu wurde die Plattform Android und der Reader Moon+ gewählt. Die nachfolgenden Screenshots zeigen die Möglichkeiten von OPDS.

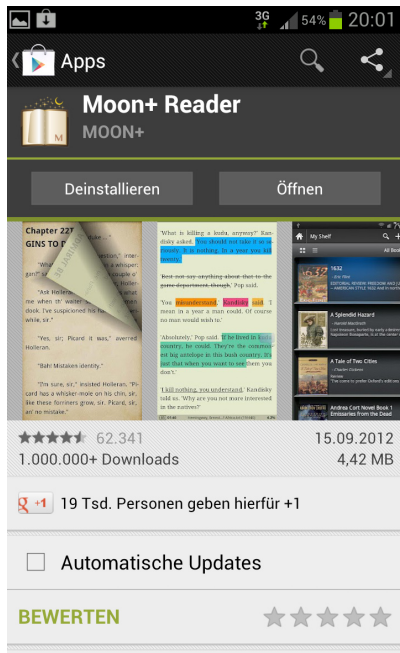


Abbildung 3.19: Moon+ im Google Playstore

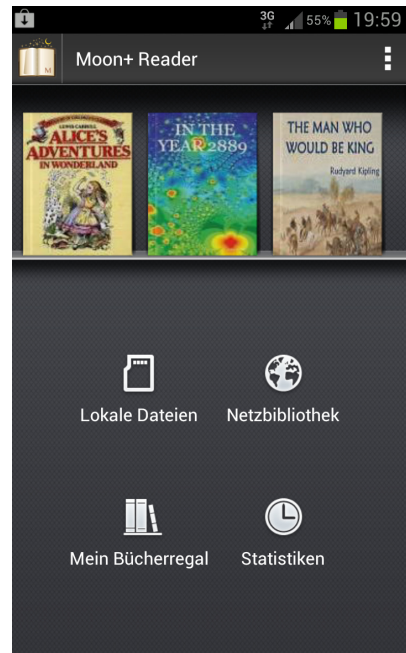


Abbildung 3.20: Moon+ Erster Start

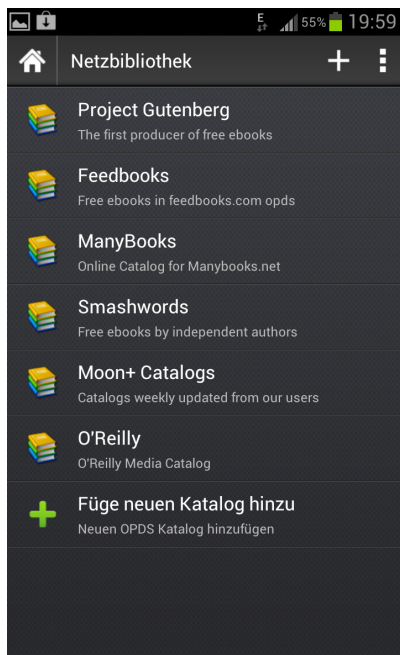


Abbildung 3.21: Hinzufügen von O'Reilly OPDS

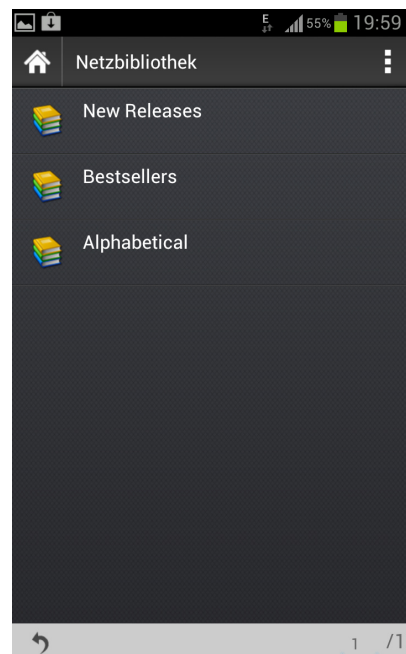


Abbildung 3.22: O'Reilly OPDS - Navigation

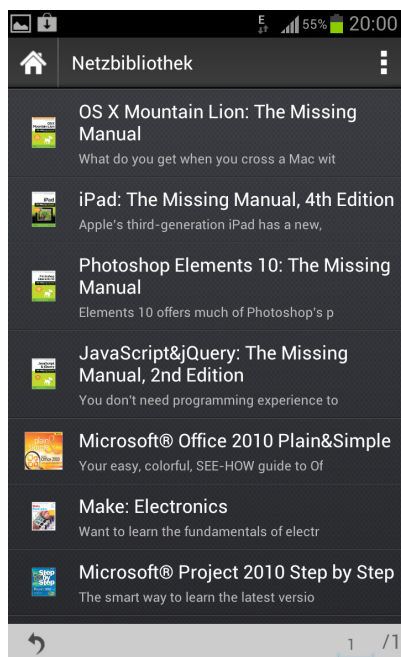


Abbildung 3.23: O'Reilly - Bestseller

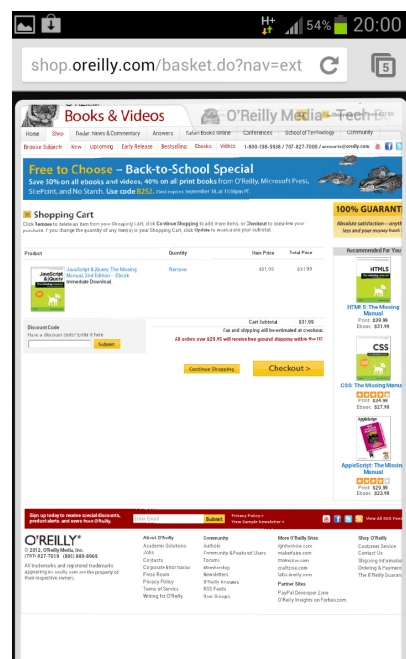


Abbildung 3.24: Weiterleitung auf O'Reilly Bookstore

Effective systems generate positive feelings of success, competence, mastery and clarity in the user community.

4 Konzepte und Systeme (B. Shneiderman 1998)

Da in den letzten Jahren die Anzahl elektronischer Publikationen massiv gestiegen ist, haben einige Institutionen und Universitäten damit begonnen Lösungen und Systeme zu entwickeln, um den neuen Anforderungen damit gerecht zu werden. All diese Systeme bauen auf einer eigenen Architektur auf und bieten unterschiedliche Services an. Es wurden weiters unterschiedliche Technologien verwendet, um diese Systeme umzusetzen. Dieses Kapitel versucht diese Systeme nach vorher festgelegten Kriterien zu bewerten. Die Auswahl der Systeme beschränkt sich hierbei auf OpenSource Produkte, die zum Großteil von Universitäten selbst entwickelt wurden. Es folgt zuerst eine kurze Erklärung der einzelnen Systeme, um danach genauer auf die vorab definierten Parameter und Kriterien eingehen zu können.

4.1 Kriterien bei der Auswahl einer digitalen Bibliothek im Bereich Universität

Die Anforderungen im Umfeld einer Universität oder Forschungseinrichtung sind meist sehr unterschiedlich. Je nachdem, wo welche Schwerpunkte gesetzt werden, kann die Auswahl sehr differenziert erfolgen. Dies bestätigen auch die hier betrachteten Systeme. Obwohl von Universitäten selbst entwickelt, gibt es doch große Unterschiede im Aufbau der Architektur oder im Umfang der angebotenen Services. [PN07]

Aufbau/Struktur des digitalen Objekts Wie wird das digitale Objekt im System gespeichert? Wie ist es aufgebaut und wie erhält man darauf Zugriff?

Möglichkeiten der Strukturierung / Relationen Welche Möglichkeiten der Kategorisierung sind gegeben? Wie können Relationen zwischen Objekten und Gruppen aufgebaut werden?

Speicherung digitaler Inhalte / Metadaten Wie werden die Objekte abgelegt? Welche Metadaten und Metadatenformate werden unterstützt? Ist der volle Umfang von Metadatenstandards gewährleistet?

Suchen und Browsen Wie werden die Objekte indiziert und wie gefunden? Welche Arten von Suchkriterien und Suchmasken werden angeboten?

Rechtemanagement und Authentifizierung Welche Methoden des Usermanagements sind vorhanden? Wie detailliert erfolgt die Zugangskontrolle? Welche Möglichkeiten der Authentifizierung sind gegeben?

Schnittstellen und Interaktionsmöglichkeiten Gibt es Möglichkeiten Daten zu exportieren oder zu importieren? Werden Schnittstellen zu anderen Systemen angeboten? Wie einfach lässt sich die digitale Bibliothek in bestehende Softwareumgebungen einbinden?

Customizingmöglichkeiten / Erweiterbarkeit Welche Customizingmöglichkeiten sind gegeben? Können Workflows zur z.B. Speicherung und Metadatenversorgung manuell abgeändert werden?

Verwendete Technologien und Services In welcher Entwicklungsumgebung wurde die Software programmiert? Welche Sprache wird verwendet und welche Services und Zusatzdienste werden angeboten?

Größe der Community Wie groß und aktiv ist die dahinter stehende Community? Gibt es eine Roadmap oder Pläne für weitere Entwicklungen?

Hardware und Softwarevoraussetzungen Welche Voraussetzungen sind von Seiten der Hard- und Software notwendig, um die digitale Bibliothek installieren zu können (Hostbetriebssystem, Softwarebibliotheken, Entwicklungssysteme, usw,...).

Die Information für die Beantwortung oben genannter Fragen wurde, wenn nicht anders gekennzeichnet, direkt vom Webauftritt der Anbieter bzw. Hoster bezogen. Fedora [[Fed12](#)] DSpace [[Fou12a](#)]

4.2 Fedora

Fedora (Flexible Extensible Digital Object Repository Architecture) wurde ursprünglich von Forschern der Cornell Universität entwickelt. Mittlerweile ist das Fedora Repository Project und die 2008 gegründete Fedora Commons Community unter der Verwaltung von DuraSpace [Fed12]. Dieses unabhängige non-profit Unternehmen ist für die Betreuung und Weiterentwicklung von Fedora verantwortlich. [Dur12] Fedora ist eine Umgebung zum Speichern, Verwalten und zur Zugangskontrolle digitaler Objekte.

4.2.1 Aufbau/Struktur des digitalen Objekts

Die kleinste Einheit bei Fedora ist das digitale Objekt. Das Fedora Objekt Modell wird mittels FOXML beschrieben. FOXML ist ein eigenes für Fedora entwickeltes Format. Es kann zum Export oder Import von Objekten verwendet werden. Das Objekt selbst besteht aus mehreren Komponenten, wie einer fixen ID, diversen Eigenschaften des Objekts und Datastreams.

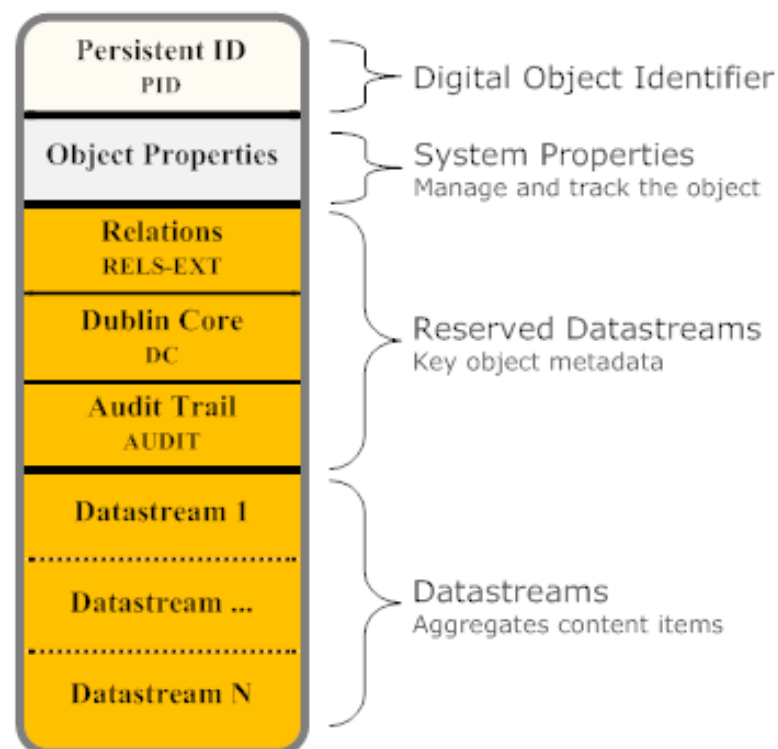


Abbildung 4.1: Fedora Objekt Modell [DW11]

Jedes Objekt in Fedora wird mit einer eindeutigen ID versehen, mit der es jederzeit eindeutig im Repository auffindbar ist. Diese ID ist persistent im System gespeichert. In den Objekteigenschaften werden notwendige Informationen gespeichert, um das Objekt verwalten zu können. Diese Information dient auch der Nachvollziehbarkeit. Datastreams beschreiben den Inhalt des Objekts. Jeder Datastream wird selbst wieder mit mehreren Eigenschaften beschrieben. Der Datastream Identifier ist nur im Umfeld des Objekts eindeutig. Dieser muss nicht im gesamten Repository eindeutig sein. In Fedora sind vier Datastream Identifier reserviert. [DW11]

Identifier	Beschreibung
DC	Jedes Fedora Objekt besitzt einen Dublin Core Datastream der Metadaten über das Objekt enthält. Werden keine Metadaten zur Verfügung gestellt, so wird dieser Stream vom System mit den DC Elementen dc:title und dc:identifier automatisch erzeugt.
AUDIT	Hier werden jegliche Änderungen am Objekt gespeichert. Dieser Datastream ist nicht bearbeitbar, da alle Änderungen selbst vom System verwaltet werden.
RELS-EXT	In diesem Bereich werden mögliche Relationen zu anderen Objekten gespeichert.
RELS-INT	In diesem Bereich werden interne Relationen im Objekt selbst gespeichert. Es kann dabei auch auf eigens definierte Datastreams im Objekt selbst referenziert werden.

Tabelle 4.1: Reservierte Datastream Identifier eines Fedora Objekts [DW11]

4.2.2 Möglichkeiten der Strukturierung / Relationen

Fedora bietet die Möglichkeit mittels dem RELS-EXT Datastream Relationen aufzubauen. Diese Objekt zu Objekt Metadaten werden mit Hilfe des Resource Description Framework (RDF) in XML gespeichert. Fedora bietet eine einfache Ontologie mit denen die Relationen isMemberOfCollection und isPartOf zwischen den Objekten ausgedrückt werden kann.

4.2.3 Speicherung digitaler Inhalte / Metadaten

In Fedora gibt es drei Möglichkeiten digitale Inhalte in Form von Datastreams zu speichern.

- intern als XML File
- direkt am Filesystem
- auf einer externen Quelle

Zusätzliche Information wie z.B. Format des digitalen Objekts werden im Datastream selbst gespeichert. Dazu gibt es ein Set an einfachen Metadaten.

4.2.4 Suchen und Browsen

Suchen ist in allen Feldern möglich, die indiziert wurden. Standardmäßig sind alle Dublin Core Metadaten und die Objekteigenschaften indiziert. Einzelne Felder können mittels boolescher Logik angesprochen und für die Suche kombiniert werden. Mit dem Webinterface von Fedora sind damit einfache Queries oder Textfeldsuchen möglich. Fedora selbst bietet keine Oberfläche zum Browsen von Inhalten an. Dazu gibt es zusätzliche Applikationen die diese Funktionalität abdecken.

4.2.5 Rechtemanagement und Authentifizierung

Fedora erlaubt Sicherheitseinstellungen, die bis auf die Tiefe von Objekten reichen. Weiters können gruppenspezifische Einstellungen vorgenommen werden. Diese Einstellungen können mittels XACML (eXtensible Access Control Markup Language) oder FeSL (Fedora Security Layer) durchgeführt werden. Die Authentifizierung kann weiters über LDAP (Lightweight Directory Access Protocol) oder definierte IP Adress Räume abgedeckt werden.

4.2.6 Schnittstellen und Interaktionsmöglichkeiten

Fedora unterstützt das Protokoll OAI-PMH, um Metadaten mit anderen Repositories austauschen zu können. Weiters können die digitalen Objekte in Fedora mittels METS XML exportiert werden. Da Fedora eindeutige und persistente Identifier für seine Objekte verwendet, haben auch externe Systeme die Möglichkeit darauf zu verweisen.

4.2.7 Customizingmöglichkeiten / Erweiterbarkeit

Fedora bietet eine hohe Flexibilität was die digitalen Inhalte betrifft. Es können mehrere Arten von Zugriff und Bearbeitung festgelegt werden. Fedora bietet auch mehrere APIs für den Zugriff und weiterer Bearbeitung. Diese sind API-A (Access), API-M (Management) und eine REST (Representational State Transfer) API.

4.2.8 Verwendete Technologien und Services

OAI-PMH Provider, SOAP, REST, FOXML,

4.2.9 Größe der Community

Es existiert ein aktiver Twitterkanal und eine Mailinglist. Es gibt eine große Community mit derzeit ca. 305 Instanzen die Fedora einsetzen.

4.2.10 Hardware und Softwarevoraussetzungen

Voraussetzung:

Java SE Development Kit (JDK) 6, Tomcat 6.0.20, Servlet 2.5/JSP 2.1, Maven2, Derby SQL Database 10.5.3.

optional:

Jetty, JBoss, MySQL, Oracle, PostgreSQL oder Microsoft SQL Server

4.3 DSpace

Die Entwicklung von DSpace begann im Jahr 2000 unter Kooperation mit HP Labs und dem MIT (Massachusetts Institute of Technology). DSpace erlaubt leichten und schnellen Zugang zu digitalen Ressourcen und erfreut sich einer großen Community. Mittlerweile wird DSpace analog wie Fedora vom Unternehmen Duraspace weiter betreut. [Fou12a]

4.3.1 Aufbau/Struktur des digitalen Objekts

Die grundlegende Organisation der Objekte ist wie folgt aufgebaut:

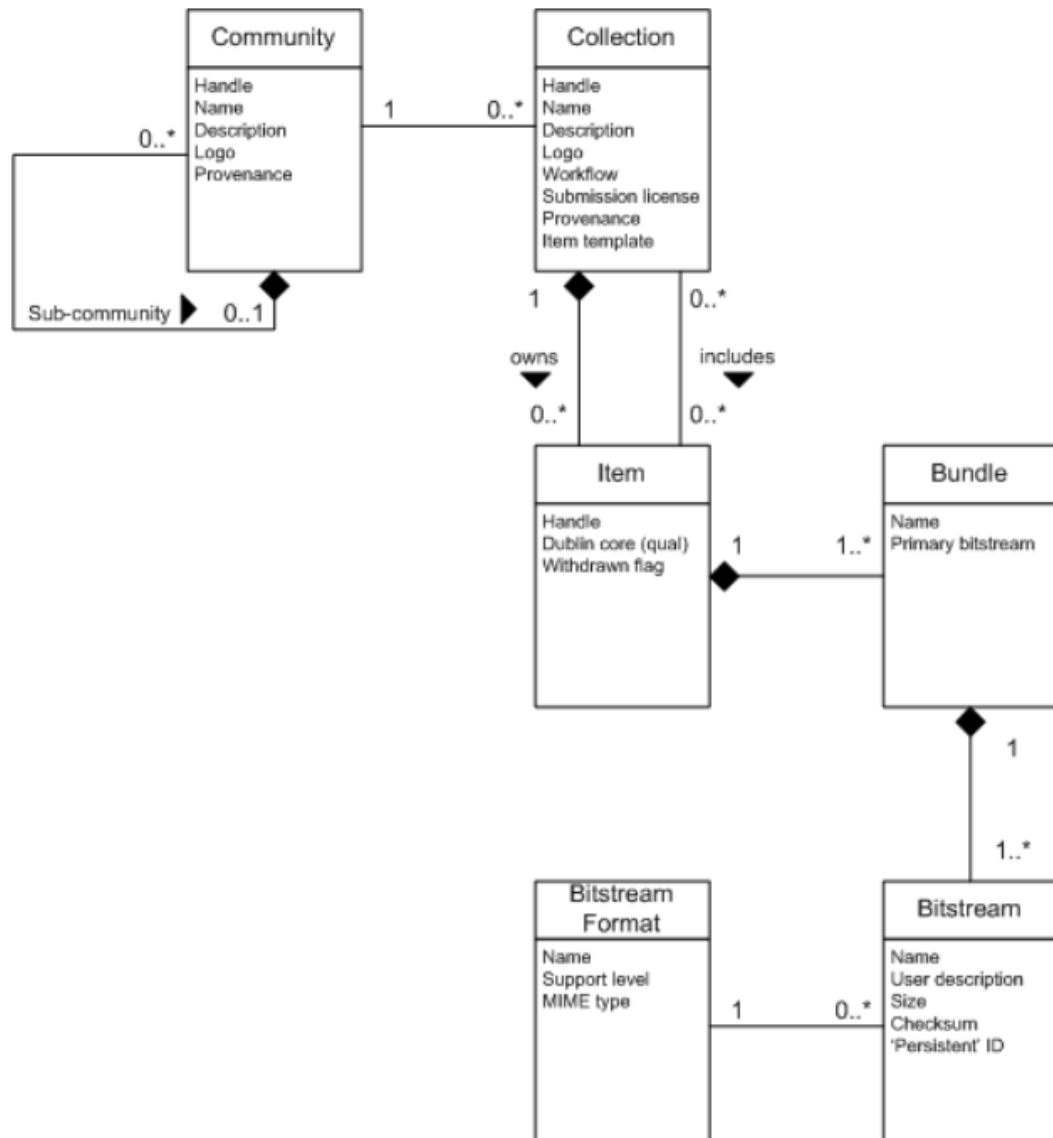


Abbildung 4.2: DSpace Daten Modell [Dig09]

DSpace versucht die lokale Struktur der Anwender abzubilden und generiert daraus eine entsprechende Systemstruktur. An einer Universität hätte somit jede Fakultät oder jedes Institut einen eigenen Bereich für deren Dokumente. Zusätzlich können auch die Verwaltung und Berechtigung gruppenspezifisch erfolgen.

Das Basiselement in DSpace nennt sich Item. In diesem Element wird Metainformation mittels Dublin Core und digitale Inhalte abgelegt. Dspace verwendet eindeutige Identifikatoren für Objekte. Diese eindeutigen Identifikatoren werden auch für jeden Bitstream verwendet. [Dig09]

Objekt	Beispiel
Community	Institut für Informationssysteme und Computer Medien
Collection	Masterarbeiten
Item	Eine spezielle Masterarbeit eines Studenten
Bundle	Eine PDF Datei und gegebenenfalls zusätzliche Information wie Bilder oder Code
Bitstream	Eine einzelne PDF Datei oder ein Quelltext eines Programms
Bitstream Format	PDF

Tabelle 4.2: Beispielunterteilung eines DSpace Objektes [Dig09]

4.3.2 Möglichkeiten der Strukturierung / Relationen

Wie bereits gezeigt, siehe Abbildung 4.2 auf Seite 89, gibt es in DSpace die Möglichkeit Collections von Items und Communities zu erstellen. Ein Item kann in einer oder mehreren Collections vorhanden sein. Es gibt derzeit keine Möglichkeit Beziehungen zwischen Items herzustellen.

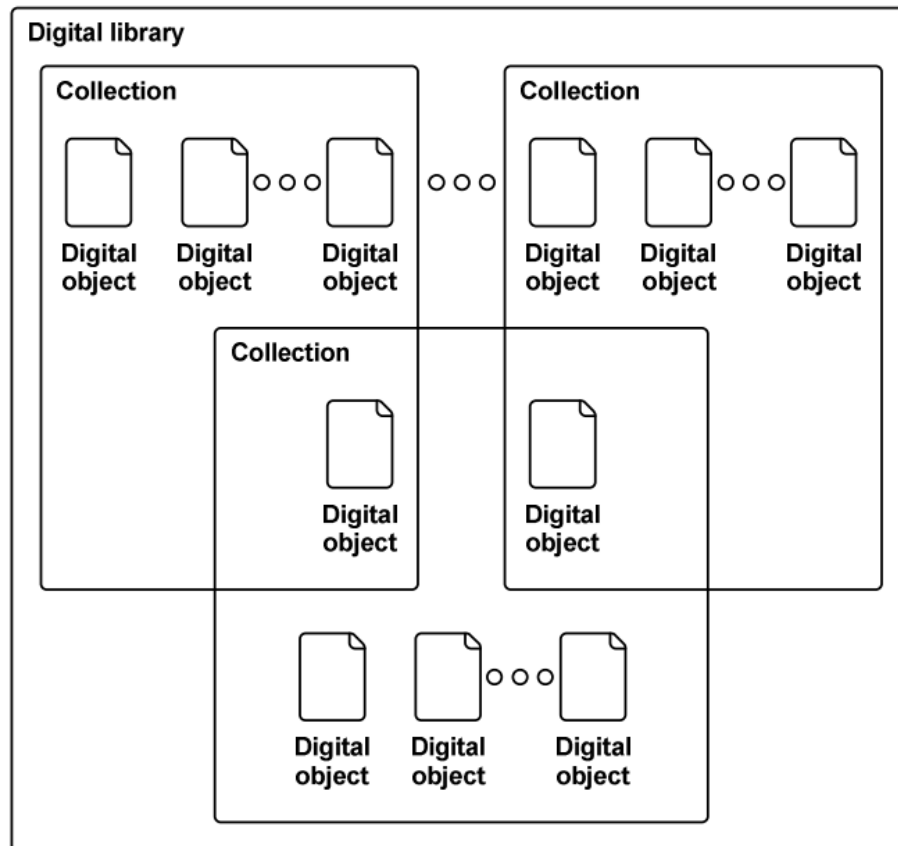


Abbildung 4.3: DSpace - Speicherung der Objekte [Kök11]

4.3.3 Speicherung digitaler Inhalte / Metadaten

DSpace speichert die Metadaten in einer relationalen Datenbank wie z.B. Oracle. Andere Information oder digitale Inhalte sind als Bitstreams vorhanden und werden in einem Filesystem abgelegt. Jeder Bitstream entspricht dabei einem Bitstream Format. In DSpace wird Metainformation mit Hilfe von Dublin Core beschrieben. Weitere beschreibende Metainformation ist möglich, nur wird diese direkt im oder mit dem Bitstream abgelegt.

4.3.4 Suchen und Browsen

Die Metainformation kann indiziert und in der relationalen Datenbank gesucht werden. Mit Hilfe der Jakarta Lucene API kann auch anders definierte Metainformation idiziert und gesucht werden. Die Suche selbst kann auf eine Community oder auf Collections eingeschränkt werden. Zum Browsen werden standardmäßig die Felder Titel, Autor und Datum verwendet.

4.3.5 Rechtemanagement und Authentifizierung

Durch die Unterteilung in Collections kann auch die Rechtevergabe bezogen auf diese vergeben werden. Grundsätzlich kann die Berechtigung auf Itemebene erfolgen wo für jeden Bitstream Lese- und Schreibberechtigungen vergeben werden können. Die Authentifizierung der Nutzer kann entweder über Passwörter, Zertifikaten (X509) oder über LDAP erfolgen. Die Berechtigungen können über ein Webinterface bearbeitet und vergeben werden.

4.3.6 Schnittstellen und Interaktionsmöglichkeiten

DSpace unterstützt OAI-PMH. Analog zu Fedora kann auch DSpace Objekte mittels METS XML exportieren. DSpace verwendet persistente Identifier für seine Objekte und bietet damit auch Zugriff für externen Systeme.

4.3.7 Customizingmöglichkeiten / Erweiterbarkeit

Das Datenmodell von DSpace ist sehr flexibel. Durch das Ablegen der Metainformation in relationalen Datenbanken ist auch eine einfache Verwaltung gegeben. Das Userinterface hingegen ist sehr fix ausgelegt und bietet nur kleine Möglichkeiten der Modifikation.

4.3.8 Verwendete Technologien und Services

XML, METS, OAI-PMH Provider, OpenURL

4.3.9 Größe der Community

Derzeit 1314 Installationen, mailinglist und aktive große Community.

4.3.10 Hardware und Softwarevoraussetzungen

Voraussetzung:

Unix-Betriebssystem Java 1.4 (Standard SDK ist ausreichend) Apache Ant 1.6.2 oder höhere Version Datenbank: PostgreSQL 7.3 (oder höher), oder Oracle 9 Servlet Container: Jakarta Tomcat 4.x/5.x

optional:

Jetty oder Caucho Resin

4.4 Fazit

Es gibt mittlerweile eine große Anzahl von Repository Anbietern. Wenn man deren Entwicklungsgeschichte betrachtet, findet sich oft ein gemeinsamer Nenner. Der Ursprung findet sich oft in universitärer Umgebung. Gerade Universitäten und Forschungseinrichtungen sind immer auf der Suche nach Lösungen, um große Mengen an Information sinnvoll ablegen und speichern zu können. Viele dieser Systeme sind daher aus einer zuerst lokal begrenzten Entwicklung entstanden. Daher gibt es auch Systeme für fast jedes Betriebssystem und unterschiedlichen Technologien. Die Anschaffung einer solchen Lösung wird sich in erster Linie an den lokalen Gegebenheiten und der notwendigen Strukturen richten. Sollten größere Anpassungen vorgenommen werden ist vielleicht auch der Schwerpunkt der Programmiersprachen Voraussetzung. [Dobo7]

4.4.1 Weitere Anbieter von Repository Systemen

Opus, EPrints, MyCoRe, ARNO, CDSware, i-TOR, Greenstone, Keystone, CDS Invenio, Digital Commons,...



Abbildung 4.4: Repository Systeme [Ber11]

A well-designed and humane interface does not need to be split into beginner and expert subsystems.

(Jef Raskin 2000)

5 Erweiterungen

Zusätzlich zu den bereits bestehenden Features der vorhin besprochenen Systeme gibt es weitere Möglichkeiten deren Bedienung und Nutzung durch Weiterentwicklungen komfortabler zu gestalten. In diesem Kapitel werden teils praktische Umsetzungen von zusätzlichen Features beschrieben und erklärt.

5.1 Online EPUB eBook Reader

Eine praktische Erweiterung in der Desktop Umgebung wäre es einen EPUB Reader ausschließlich im Browser zu realisieren. Dieser sollte ohne Browserplugin oder anderen zusätzlichen Hilfsmitteln funktionieren. Die verwendeten Technologien sollten dabei nur HTML, Javascript und CSS sein.

Zur Entwicklung wurde unter anderem das JavaScript Framework jQuery verwendet. Dieses Framework bietet umfangreiche Klassenbibliotheken und Funktionen zur DOM-Manipulation. jQuery ist frei und wird laufend weiter entwickelt. Zusätzlich bietet die Community um jQuery eine umfangreiche Sammlung von Plugins an, die auch kostenlos und frei verwendet werden können. [Fou12b]

5.1.1 Erstellung der eReader Oberfläche

Als erster Schritt der Entwicklung wurde der grundlegende Aufbau des Readers überlegt. Dazu wurde das Layout mit mehreren Divs erstellt und unterteilt.

Div Name	Beschreibung
navleft	beinhaltet das Inhaltsverzeichnis des geladenen eBooks
content	beinhaltet die einzelnen Kapitel des geladenen eBooks
pagecountdiv	beinhaltet die Bereiche in denen die Seitenanzahl angezeigt wird
leftpagecountdiv	beinhaltet die berechnete Seitenanzahl links
rightpagecountdiv	beinhaltet die berechnete Seitenanzahl rechts
navbottom	beinhaltet die beiden Buttons mit denen durch den Inhalt des Buches navigiert werden kann
epub	beinhaltet mehrere EPUB eBooks mit unterschiedlichen Eigenschaften, die in den online EPUB Reader geladen werden können

Tabelle 5.1: Funktion und Beschreibung der verwendeten Divs

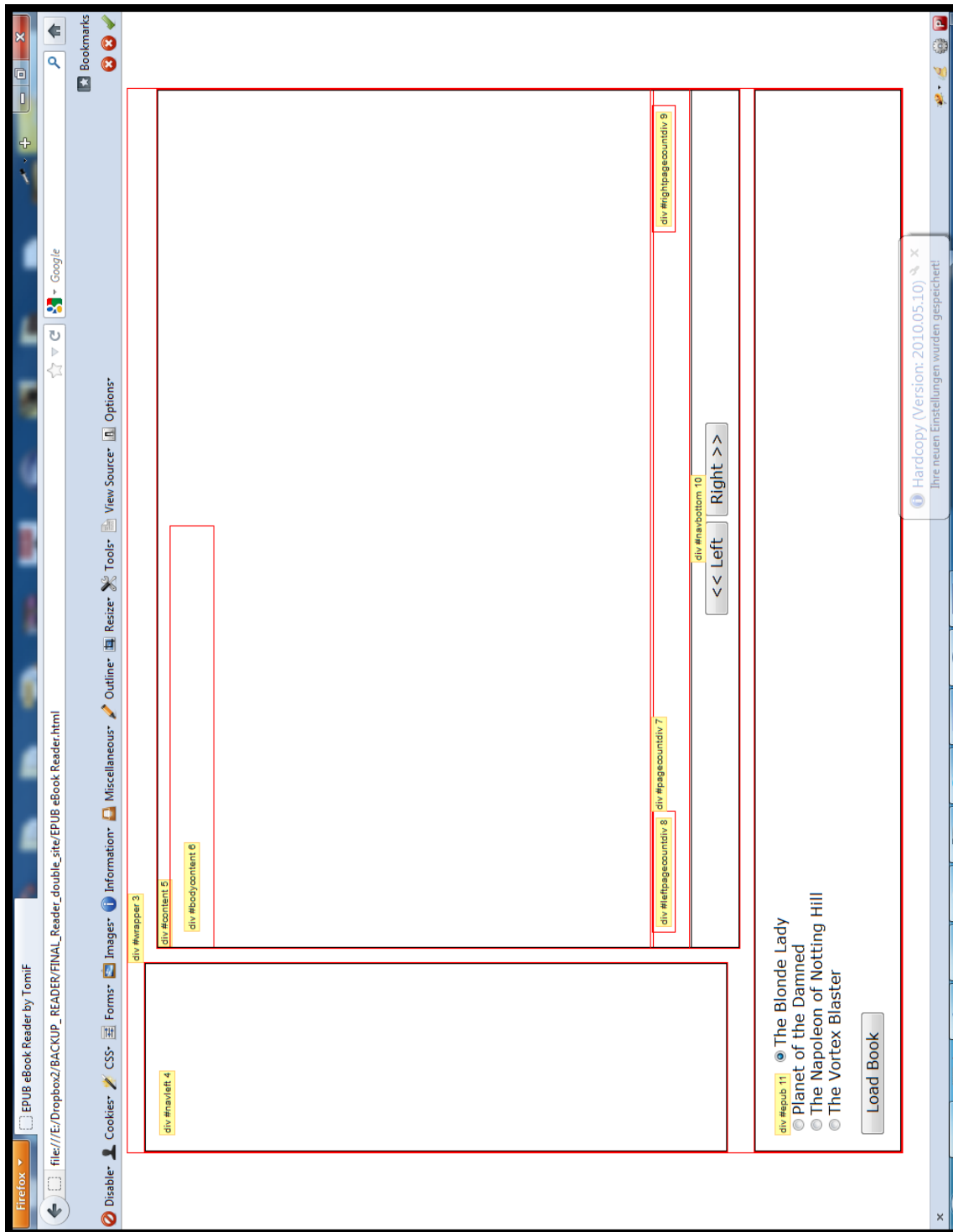


Abbildung 5.1: Unterteilung des Layouts mit Hilfe von Divs

5.1.2 Parsen der *.ncx Datei

Einer der ersten Schritte war das Auslesen und Darstellen des Inhalts. Das Inhaltsverzeichnis wird mittels der Datei *.ncx dargestellt. Um die Datei richtig interpretieren zu können, muss die Bedeutung der einzelnen Tags klar definiert sein.

```

<?xml version="1.0" encoding="UTF-8"?>
<navMap>
  <navPoint id="navpoint-1" playOrder="1">
    <navLabel>
      <text>Title Page</text>
    </navLabel>
    <content src="title.xml"/>
  </navPoint>
  <navPoint id="navpoint-2" playOrder="2">
    <navLabel>
      <text>epubBooks Information</text>
    </navLabel>
    <content src="epubbooksinfo.xml"/>
  </navPoint>
  <navPoint id="navpoint-3" playOrder="3">
    <navLabel>
      <text>FIRST EPISODE</text>
    </navLabel>
    <content src="book-01.xml"/>
  <navPoint id="navpoint-4" playOrder="4">
    <navLabel>
      <text>CHAPTER I</text>
    </navLabel>
    <content src="chapter-001.xml"/>
  </navPoint>
  <navPoint id="navpoint-5" playOrder="5">
    <navLabel>
      <text>CHAPTER II</text>
    </navLabel>
    <content src="chapter-002.xml"/>
  </navPoint>
  .....
  <navPoint id="navpoint-9" playOrder="9">
    <navLabel>
      <text>CHAPTER VI</text>
    </navLabel>
    <content src="chapter-006.xml"/>
  </navPoint>
  .....
  <navPoint id="navpoint-13" playOrder="13">
    <navLabel>
      <text>Footnotes</text>
    </navLabel>
    <content src="footnotes.xml"/>
  </navPoint>
</navMap>
</ncx>

```

Quellcode 5.1: Ausschnitt einer *.ncx Datei

Jeder Abschnitt, der mit einem <navPoint> Tag markiert ist, identifiziert einen Bereich im Dokument, zu dem der Nutzer direkt navigieren kann. Das Ele-

ment `<navPoint>` besitzt weiters die zwei Attribute »id« und »playOrder«. Mit dem Attribut »id« wird eine im Dokument eindeutige ID vergeben. Das Attribut »playOrder« definiert die Reihenfolge der Elemente im Inhaltsverzeichnis.

Das Element `<navPoint>` besteht aus den Elementen `<navLabel>` und `<content>`. Das Element `<navLabel>` besitzt auch ein Element `<text>`. Der Text, der sich im Element `<text>` befindet, wird in weiterer Folge als Teil der Navigation ausgegeben. z.B.:

```
<text>Title Page</text>
```

Das Element `<content>` verweist dabei auf die Datei, in der der Inhalt des Abschnitts gespeichert ist.

```
<content src="title.xml"/>
```

Mit dieser Information lässt sich die Datei *.ncx parsen und eine Navigation des gesamten Inhalts erstellen. Diese wird im Div »navleft« angezeigt.

5.1.3 Laden und Anzeigen der einzelnen Kapitel

Der nächste Schritt der Entwicklung war die Anzeige des Inhalts. Die Referenz zu den Inhalten ist über die Navigation gegeben. Unterer Codeabschnitt zeigt eine verkürzte Version einer XHTML Datei.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html
PUBLIC "-//W3C//DTD XHTML 1.1//EN" "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en">
  <head>
    <title>CHAPTER I – NUMBER 514, SERIES 23 | The Blonde Lady</title>
    <link rel="stylesheet" href="css/book.css" type="text/css"/>
    <meta http-equiv="Content-Type" content="application/xhtml+xml; charset=utf-8"/>
    <meta name="EPB-UUID" content="A1989627-6C07-1014-9E83-850FB41434CC"/>
  </head>
  <body>
    <div class="body">
      <div class="chapter">
        <h3 class="chapter-title">CHAPTER I</h3>
        <h4 class="chapter-subtitle">NUMBER 514, SERIES 23</h4>
        <p>On the 8th of December last, M.&nbsp;Gerbois, professor of mathematics at Versailles College, rummaging among the stores at a second-hand dealer's, discovered a small mahogany writing desk, which took his fancy because of its many drawers.</p>
        <p>"That's just what I want for Suzanne's birthday," he thought.</p>
```

```

<p>M. Gerbois' means were limited and, anxious as he was to please his daughter, he felt it
  his duty to beat the dealer down. He ended by paying sixty&#8211;five francs. As he was
  writing down his address, a well&#8211;groomed and well&#8211;dressed young man,
  who had been hunting through the shop in every direction, caught sight of the writing
  &#8211;desk and asked:</p>
<p>"How much for this?"</p>
<p>"It&#8217;s sold," replied the dealer.</p>
<p>"Oh&#8230; to this gentleman?"</p>
<p>M. Gerbois bowed and, feeling all the happier that one of his fellow&#8211;men envied
  him his purchase, left the shop. But he had not taken ten steps in the street before the
  young man caught him up and, raising his hat, said, very politely:</p>
<p>But M.&nbsp;Gerbois remained convinced that he had suffered an enormous loss. A
  fortune must have been concealed in some secret drawer and that was why the young
  man, knowing of the hiding&#8211;place, had acted with such decision.</p>
<p>"Poor father! What should we have done with the fortune?" Suzanne kept saying.</p>
<p>"What! Why, with that for your dowry, you could have made the finest match going!"</p>
>
<p>Suzanne aimed at no one higher than her cousin Philippe, who had not a penny to bless
  himself with, and she gave a bitter sigh. And life in the little house at Versailles went on
  gaily, less carelessly than before, shadowed over as it now was with regret and
  disappointment.</p>
<p class="thoughtbreak">* * * * *</p>
<p>Two months elapsed. And suddenly, one after the other, came a sequence of the most
  serious events, forming a surprising run of alternate luck and misfortune.</p>
<p>On the 1st of February, at half&#8211;past five, M.&nbsp;Gerbois, who had just come
  home, with an evening paper in his hand, sat down, put on his spectacles and began to
  read. The political news was uninteresting. He turned the page and a paragraph at once
  caught his eye, headed:</p>
<blockquote>
  <div class="line-group">"THIRD DRAWING OF THE PRESS&#8211;ASSOCIATION
    LOTTERY"</div>
  <div class="line-group">"First prize, 1,000,000 francs: No.&nbsp;514, Series 23."</div>
</blockquote>
<p>The paper dropped from his hands. The walls swam before his eyes and his heart
  stopped beating. Number 514, series 23, was the number of his ticket! He had bought it
  by accident, to oblige one of his friends, for he did not believe in luck; and now he had
  won!</p>
<p>He took out his memorandum&#8211;book, quick! He was quite right: number 514, series
  23, was jotted down on the fly&#8211;leaf. But where was the ticket?</p>
  . . . . .
<p>"Once there&#8217;s no trace of their flight, they must be here!"</p>
<p>It is possible that, in the depths of his mind, he was less firmly convinced. But he refused
  to admit as much to himself. No, a thousand times no: a man and a woman do not
  vanish into space like the wicked genii in the fairy&#8211;tales! And, without losing
  courage, he continued his searchings and investigations, as though he hoped to discover
  them hidden in some impenetrable retreat, bricked up in the walls of the house.</p>
</div>
</div>
</body>
</html>

```

Quellcode 5.2: Auszug einer XHTML Datei

Das Laden eines dieser XHTML Dateien in den Browser ist relativ einfach. Trotzdem gab es bei der Anzeige einige Probleme zu lösen. Eines der Probleme war, den Inhalt selbst im Bereich des Divs anzuzeigen und abzugrenzen. Der erste Versuch, den Inhalt in den dafür vorgesehenen Div Bereich zu laden, ergab folgende Sicht.

The screenshot shows a Firefox browser window displaying an EPUB eBook reader. The address bar shows the file path: file:///E:/Dropbox/BACKUP_READERS/FINAL_Reader_double_site/EPUB_eBook_Reader.html. The browser's toolbar includes various icons for navigation and settings.

The main content area is divided into three sections:

- Left Navigation:** A list of links for navigating through the book:
 - Title Page
 - epubBooks Information
 - FIRST EPISODE
 - CHAPTER I
 - CHAPTER II
 - CHAPTER III
 - CHAPTER IV
 - CHAPTER V
 - CHAPTER VI
 - SECOND EPISODE
 - CHAPTER I
 - CHAPTER II
 - Footnotes
- Center Content:** The main text of the chapter, starting with "CHAPTER I" and "NUMBER 514, SERIES 23". The text describes a scene at Versailles College involving a mathematician named M. Gerbois and a young man. It includes several lines of dialogue:
 - "That's just what I want for Suzanne's birthday," he thought.
 - M. Gerbois' means were limited and, anxious as he was to please his daughter, he felt it his duty to beat the dealer down. He ended by paying sixty-five francs. As he was writing down his address, a well-groomed and well-dressed young man, who had been hunting through the shop in every direction, caught sight of the writing-desk and asked:
 - "How much for this?"
 - "It's sold," replied the dealer.
 - "Oh... to this gentleman?"
 - M. Gerbois bowed and, feeling all the happier that one of his fellow-men envied him his purchase, left the shop. But he had not taken ten steps in the street before the young man caught him up and, raising his hat, said, very politely:
 - "I beg a thousand pardons, sir... I am going to ask you an indiscreet question... Were you looking for this desk rather than anything else?"
 - "No. I went to the shop to see if I could find a cheap set of scales for my experiments."
 - "Therefore, you do not want it very particularly?"
 - "I want it, that's all."
 - "Because it's old I suppose?"
 - "Because it's useful."
 - "In that case, would you mind exchanging it for another desk, quite as useful, but in better condition?"
 - "This one is in good condition and I see no point in exchanging it."
 - "Still..."
- Right Navigation:** A list of book titles with radio buttons next to them:
 - The Blonde Lady
 - Planet of the Damned
 - The Napoleon of Notting
 - The Vortex Blaster
 Below this list is a "Load Book" button.

At the bottom of the main content area, there is a row of navigation buttons: "Left", "Right", and ">>".

Abbildung 5.2: Text außerhalb der Div Grenzen

Die Aufgabe war es nun, den Inhalt auf die Größe des Divs anzupassen. Es wurden zwei Ansätze getestet.

1. Zählen der Paragraphen
2. Berechnung und Teilung der Gesamthöhe des Kapitels

Beide Lösungen erzielten sehr gute Ergebnisse, solange der Inhalt des Kapitels ausschließlich aus Text bestand. Probleme in der Berechnung und Teilung ergaben sich erst durch eingebettete Bilder oder andere Formatierungen wie z.B.: <quotes>. Manche Abschnitte sind weiters sehr tief verschachtelt und deren Elemente mit eigenen CSS Definitionen vom Buch selbst belegt, sodass eine genaue Berechnung der benötigten Höhe nicht vollständig möglich war. Es war z.B. im vorhinein nicht möglich die Höhen der Bilder mit einzuberechnen, sodass Text im online EPUB Reader nicht mehr vollständig angezeigt wurde oder dieser wieder über die Grenzen des Div hinaus angezeigt wurde.

Ein weiteres Problem unter Punkt 2 war, dass die Teilung der gesamten Höhe des Kapitels nicht immer 100% genau auf das Div Ende zugeschnitten war. Das Problem lässt sich durch folgende Grafik darstellen.

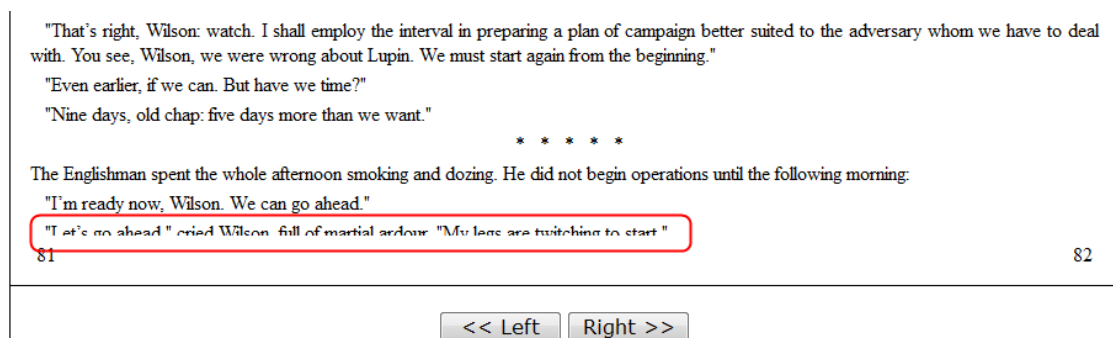


Abbildung 5.3: Text von Div Grenze geschnitten

Diese Probleme konnten durch zusätzliche Berechnungen nicht gelöst werden. Auch einige viel versprechende jQuery Plugins konnten die Problem nur teilweise lösen. Schlussendlich wurden die Probleme mit der CSS3 Eigenschaft »multi-column« gelöst. Mit dieser Eigenschaft ist es möglich beliebigen Inhalt in mehreren Spalten mit definierten Eigenschaften anzuzeigen. Eine genauere Aufschlüsselung der Formatierungsmöglichkeiten findet sich auf der Seite des W3C [W3C11].

Der Vorteil, die Formatierung des Inhalts in CSS und nicht in z.B. Tabellen vorzunehmen, ist die größere Flexibilität. Im Online EPUB Reader wurde über dem Div, welches den Inhalt des Buches anzeigt, folgendes CSS gelegt.

```
#content {  
    float: right;  
    width: 900px;  
    height: 520px;  
    margin-top: 16px;  
    border-top: 1px solid black;  
    border-left: 1px solid black;  
    border-right: 1px solid black;  
    overflow: hidden;  
    column-count: 2;  
    column-gap: 16px;  
    -moz-column-count: 2;  
    -moz-column-gap: 16px;  
    -webkit-column-count: 2;  
    -webkit-column-gap: 16px;  
    -webkit-column-rule-width: thin;  
    -webkit-column-rule-style: solid;  
    -webkit-column-rule-color: black;  
}
```

Quellcode 5.3: CSS Definition

Da zum Zeitpunkt der Entwicklung die endgültige CSS3 Spezifikation noch immer nicht als Fassung im World Wide Web Consortium (W3C) veröffentlicht wurde, verwenden einige Browserhersteller spezifische Tags, um deren Funktionalität bereits im Vorhinein abbilden zu können. Daher wurden die CSS3 Anweisungen mit dem Präfix »-moz« für Gecko basierte Browser wie Firefox oder »-webkit« für Webkit basierte Browser wie Chrome oder Safari verwendet. Dadurch entsteht auch ein kleiner Nachteil. Einige Browser unterstützen diesen Befehl gar nicht wie z.B. der Internet Explorer, andere wiederum implementieren die CSS3 Funktionalität nicht immer gleich. Dadurch können weitere Probleme in der browserübergreifenden Entwicklung entstehen.

Nach Implementierung von »multi-column« war die Ansicht des online EPUB Readers wie auf nachfolgender Grafik ersichtlich.

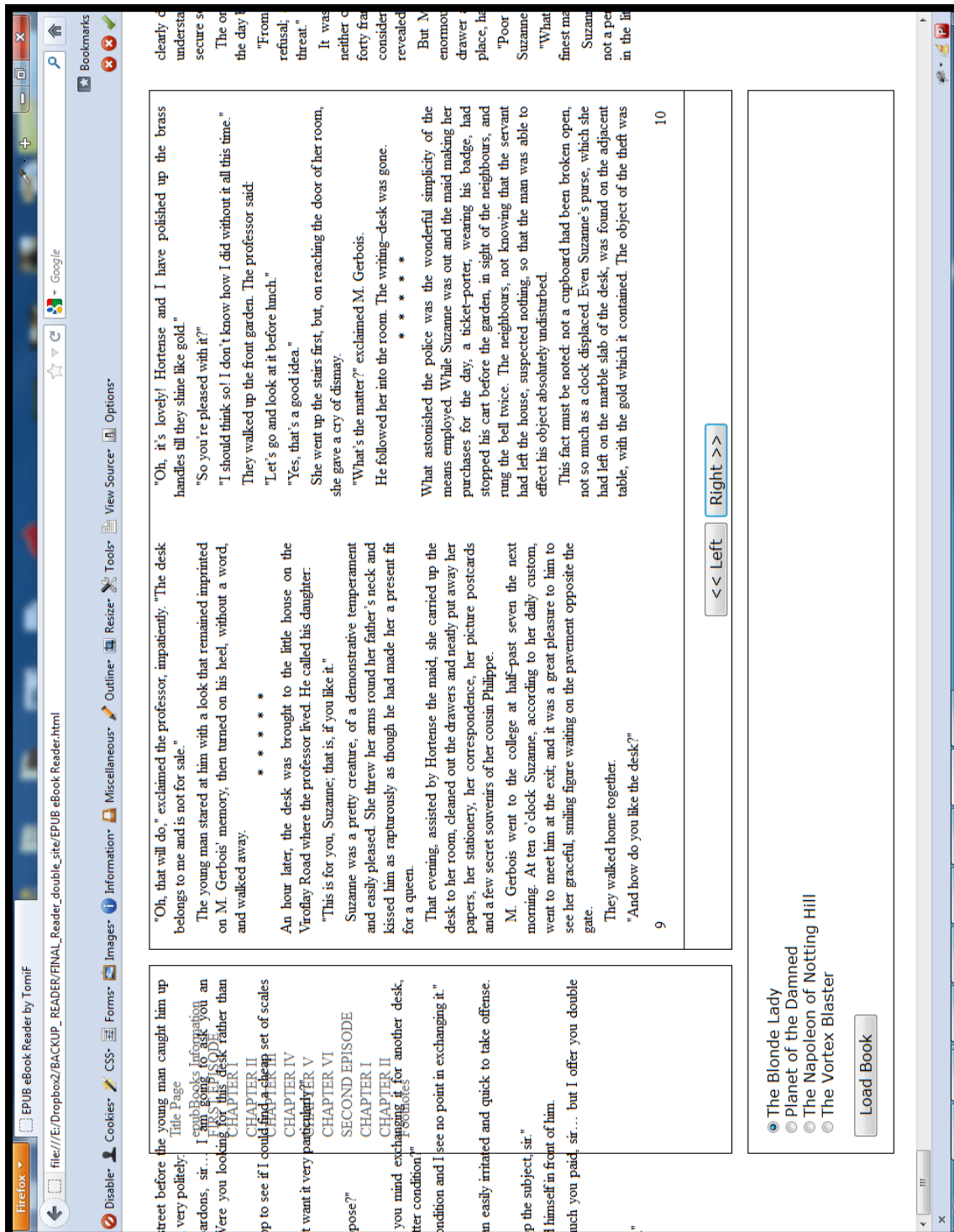


Abbildung 5.4: Mit Hilfe von multi-column geladener Text

5.1.4 Darstellung der einzelnen Seiten

Die nächste Aufgabe bestand darin, den Inhalt, der über das Inhaltsverzeichnis geladen wurde, innerhalb der Div Grenzen anzuzeigen. Dazu wurde ein weiteres Div mit dem Namen »bodycontent« innerhalb des Divs »content« angelegt. Nun wurde über CSS die Eigenschaft position des Div »bodycontent« auf relative gesetzt. Weiters wurde die Eigenschaft overflow des Div »content« auf hidden gesetzt. Dadurch wird der Inhalt, der außerhalb der Div Grenzen angezeigt wird, auf unsichtbar gesetzt.

Um nun ein »Blättern« zu simulieren, muss nun das Div »bodycontent« im Div »content« verschoben werden. Nachfolgende Grafik zeigt, wie der Inhalt verschoben wird.

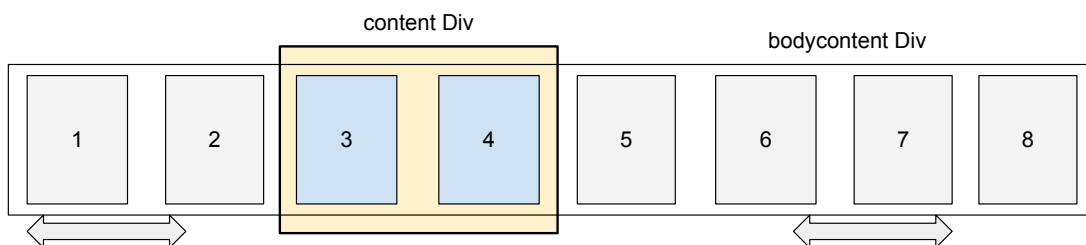


Abbildung 5.5: Verschiebung der einzelnen Spalten

Um ein Div innerhalb eines anderen verschieben zu können, wird hier die CSS Eigenschaft »right« verwendet. Mit Hilfe dieser Eigenschaft lässt sich das Div »bodycontent« um eine fixe Länge nach rechts verschieben.

Das Blättern selbst wird durch zwei Buttons ermöglicht, die unterhalb des »content« Div angezeigt werden. Diese Buttons selbst werden in einem eigenen Div »navbottom« angezeigt.



Abbildung 5.6: eReader Navigation

5.1.5 Berechnung der Seitenanzahl

Um eine korrekte Seitenanzahl zu berechnen, ist es wichtig zu wissen, wann der geladene Inhalt zu Ende ist und wann ein neues Kapitel geladen werden muss. Gleiches gilt beim Zurückblättern. Wenn es beim Zurückblättern einen Kapitelwechsel gibt, muss nicht nur der Wechsel erkannt werden, es müssen auch die letzten Seiten des vorangegangenen Kapitels angezeigt werden. Dadurch ist es notwendig zu wissen, aus wievielen Seiten, sprich Spalten die jeweiligen Kapitel bestehen. Wichtig ist auch die Einberechnung des vorab gewählten Layouts, welches durch die CSS₃ Eigenschaft »multi-column« festgelegt wurde.

Um diese Seitenberechnungen durchführen zu können, waren anfangs folgende Schritte notwendig.

1. Laden jeder Seite/jedes Kapitels in den »content« Div Container
2. Berechnen der gesamten Breite jedes einzelnen Kapitels
3. Teilen der Kapitelbreite durch die Breite des »content« Div Container
4. Speichern der Information in einem Javascript Array
5. Zählen und Speichern der jeweils letzten Seiten eines vorangegangenen Kapitels in einem weiteren Javascript Array.

Durch die Berechnung der Zahlen ist festgelegt, wie oft z.B. der Button »Right« zum Blättern pro Kapitel gedrückt werden kann, bis das Ende des Kapitels erreicht ist. Dadurch weiß man auch, dass der Inhalt nicht weiter verschoben werden darf und dass ein neues Kapitel geladen werden muss. Bei jedem Klick werden dadurch auch die Divs »leftpagecountdiv« und »rightpagecountdiv« , welche die aktuellen Seitenanzahlen anzeigen, aktualisiert.

Da die Kapitel für die Berechnung vorab in das content Div geladen wurden, sah der Nutzer alle Seiten in einem Schnelldurchlauf. Da dieser Schnelldurchlauf aller Seiten als störend empfunden werden kann, wurden zwei weitere Divs angelegt. Diese wurden mit der CSS Eigenschaft »top« auf -3000 gesetzt und als »hidden« deklariert. Dadurch wird das Vorabladen der Inhalte vor dem Nutzer verborgen. Die relevanten Divs wurden mit dem Präfix »hidden« versehen. Sie lauten daher »hiddencontent« und »hiddenbodycontent«. Sollte der Wunsch bestehen, das Vorabladen zu sehen, müssten die »hidden« Divs durch die regulären Divs ersetzt werden.

The screenshot shows a Firefox browser window displaying an eReader application. The browser's address bar shows the file path: `file:///E:/DropboxZ/BACKUP_READERS/FINAL_Reader_double_site/EPUB eBook Reader.html`. The eReader interface includes a top toolbar with icons for navigation and settings, and a bottom toolbar with buttons for navigation and a 'Load Book' button.

The main content area is divided into three sections:

- Table of Contents (Left):** A list of navigation links including 'Title Page', 'epubBooks Information', 'FIRST EPISODE', 'CHAPTER I', 'CHAPTER II', 'CHAPTER III', 'CHAPTER IV', 'CHAPTER V', 'CHAPTER VI', 'SECOND EPISODE', 'CHAPTER I', 'CHAPTER II', and 'Footnotes'.
- Main Text (Center):** The text of 'CHAPTER V KIDNAPPED'. It begins with 'Hollock Shears restrained his feelings...' and continues with a narrative about a man in a red stone. The page number '149' is displayed in a red box at the bottom right of this section.
- Text Continuation (Right):** The text continues from the previous page, starting with 'And, addressing them, he said, pointing to the door of the next house:'. It includes several lines of dialogue and narrative. The page number '150' is displayed in a red box at the bottom right of this section.

Navigation buttons '<< Left' and 'Right >>' are located between the two text sections.

The bottom right section contains a list of books with radio buttons:

- The Blonde Lady
- Planet of the Damned
- The Napoleon of Notting Hill
- The Vortex Blaster

A 'Load Book' button is positioned below this list.

Abbildung 5.7: eReader mit implementiertem Paging

5.1.6 Anbieten verschiedener EPUB eBooks

Um das Laden verschiedener EPUB eBooks zu zeigen, wurden einige Beispiele mit unterschiedlichen Eigenschaften zur Verfügung gestellt. Die eBooks wurden von der Seite www.epubebooks.com geladen. Die EPUB Dateien wurden lokal gespeichert, entpackt und im online EPUB Reader zum Test eingebunden. Jedes einzelne EPUB eBook wurde mit seiner speziellen Struktur in einem eigenen Ordner abgelegt.

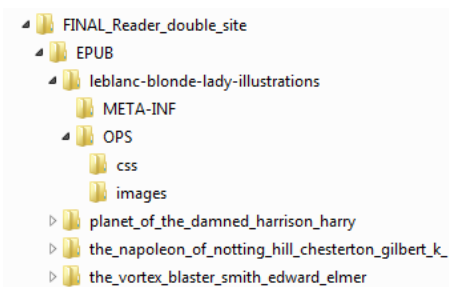


Abbildung 5.8: EPUB Ordnerstruktur

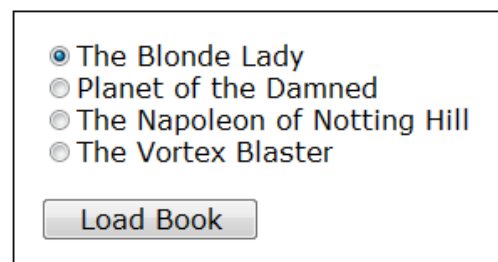


Abbildung 5.9: Karateka 1986

Durch das Testen mit unterschiedlichen Quellen wurden noch weitere Punkte gefunden. Eine Anzeige der Inhalte war nicht immer vollständig korrekt möglich. Auf diese Punkte wird in den folgenden Abschnitten eingegangen.

5.1.7 Wechsel der spezifischen EPUB CSS Dateien

Wie im Aufbau siehe Abschnitt 3.1.2 auf Seite 46 von EPUB Dateien beschrieben, besitzt jede EPUB Datei eine eigene CSS Definition. Mit dieser werden verschiedene Abschnitte des Buches spezifiziert. Jedes Buch wird per AJAX (Asynchronous JavaScript and XML) geladen. Zum Zeitpunkt des Ladens muss auch das aktuelle CSS geladen werden. Zu diesem Zweck wird ein eigenes jQuery Plugin mit dem Namen »cookie« verwendet.

Dazu ist es notwendig, den Pfad jeder CSS Datei zu wissen. Dieser wird im Div »epub« hinter den einzelnen Radio Buttons der EPUB Dateien hinterlegt.

```
<div id="epub">
  <input type="radio" name="books" id="a1"
    rel="EPUB/leblanc-blonde-lady-illustrations/OPS/css/book.css"
    value="leblanc-blonde-lady-illustrations">The Blonde Lady<br>
  <input type="radio" name="books" id="b1"
    rel="EPUB/planet_of_the_damned_harrison_harry/OPS/css/book.css"
    value="planet_of_the_damned_harrison_harry">Planet of the Damned<br>
```

```
<input type="radio" name="books" id="c1"
      rel="EPUB/the_napoleon_of_notting_hill_chesterton_gilbert_k_/OPS/css/book.css"
      value="the_napoleon_of_notting_hill_chesterton_gilbert_k_">The Napoleon of Notting
      Hill<br>
<input type="radio" name="books" id="d1"
      rel="EPUB/the_vortex_blaster_smith_edward_elmer/OPS/css/book.css"
      value="the_vortex_blaster_smith_edward_elmer">The Vortex Blaster<br>
<br>
<button id="load">Load Book</button>
</div>
```

Quellcode 5.4: Einzeln gesetzte CSS Pfade

5.1.8 Anzeige der Bilder

Bilder und Grafiken sind bei entpackten EPUB eBooks immer in einem eigenen Ordner abgelegt. Im Code der XHTML Quellen ist dieser Pfad immer relativ angegeben.

Pfad: “images/p010.jpg”

Durch das Anbieten mehrerer eBooks im online EPUB Reader wurde eine weitere Ordner Ebene erstellt, in der diese Bücher angeboten werden. Dadurch ändert sich aber auch der Pfad zu den Bildern. Wurde ein EPUB Buch direkt im Reader aufgerufen, war der Pfad korrekt.

Der neue Pfad der Bilder war nicht mehr wie in der Quelle selbst angegeben, sondern referenzierte zu den neu angelegten Ordnern. Der neue Pfad zu den Bildern eines Buches war also:

Pfad: “EPUB/the_vortex_blaster_smith_edward_elmer/OPS/images/p010.jpg”

Eine Voraussetzung des Readers war es, die bestehenden Quellen des eBooks nicht zu ändern. Aufgrund dessen wurde der Pfad der Bilder zur Laufzeit, also beim Laden selbst geändert. Die Originalquellen wurden dabei nicht modifiziert. Dazu wurde eine spezielle jQuery Funktion verwendet, die es erlaubt, mittels Callback Inhalte nachzuladen und zu modifizieren.

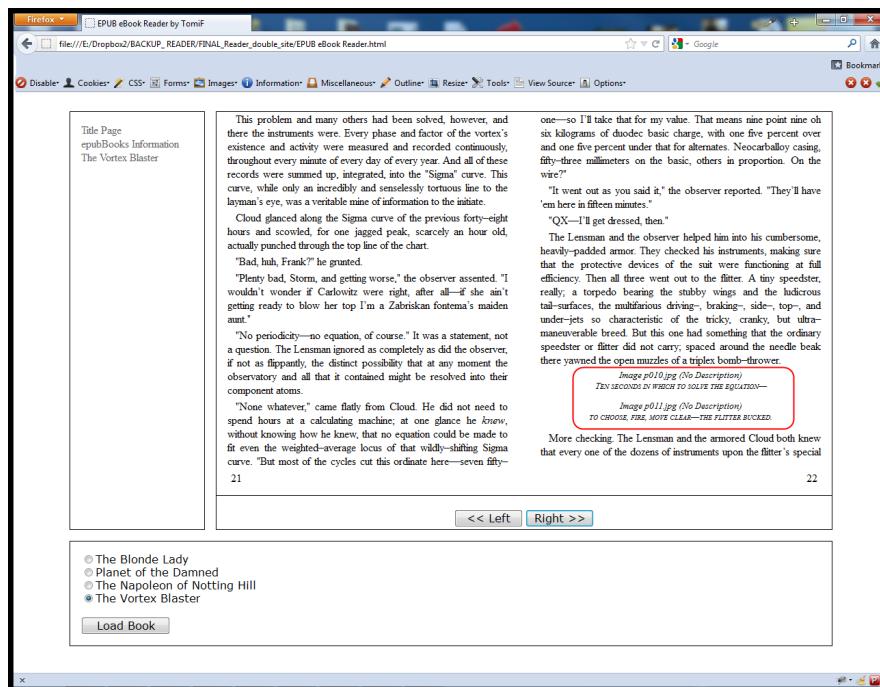


Abbildung 5.10: eReader mit falschem Pfad zu den Bildern

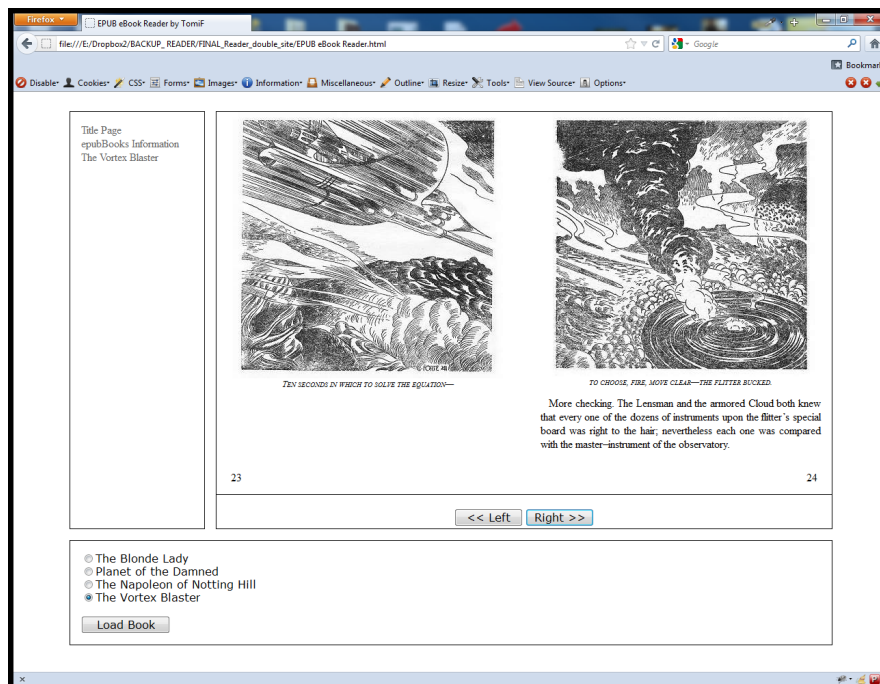


Abbildung 5.11: eReader mit richtigem Pfad zu den Bildern

5.1.9 Zweiseiten / Einseiten Layout - online EPUB Reader

Im Zuge der Entwicklung des online EPUB Readers wurde ein zweiseitiges Layout bevorzugt. Dieses kommt dem Lesen eines normalen Buches am nächsten. Da es aber auch Dokumente gibt, die sehr viel Grafik, Diagramme oder Bilder beinhalten, wurde auch eine einseitige Version entwickelt. Diese bietet mehr Platz und unterstützt auch eine Seitenangabe. Es waren dazu einige Anpassungen bei den Divs und der Ausrichtung notwendig. Die Berechnung der Seitenanzahl musste hingegen komplett überarbeitet werden.

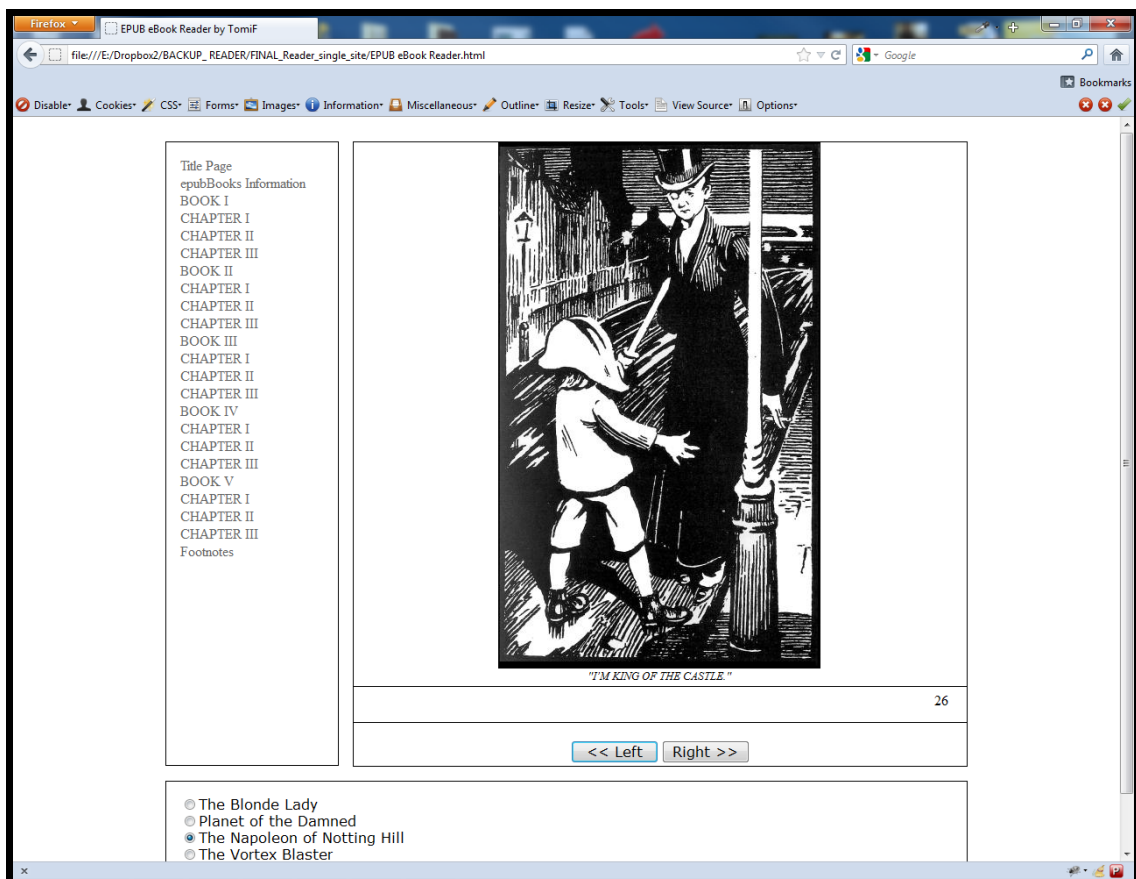


Abbildung 5.12: eReader - Layout mit einer Seite

5.1.10 Browserübergreifende Entwicklung

Zur Entwicklung des online EPUB Readers wurde hauptsächlich der Browser Mozilla Firefox gewählt. Dieser zeichnet sich dadurch aus, dass es besonders vie-

le Entwicklungstools und Erweiterungen (Add-ons) gibt. Hervorheben möchte ich dabei die Erweiterungen Firebug [Moz12] und WebDeveloper [Ped12]. Diese sind bei der Webentwicklung essentiell und waren auch hier eine sehr große Unterstützung.

Nach dem Entwicklungsprozess lief der online EPUB Reader stabil und fehlerfrei im Mozilla Firefox. Der Blick auf andere Browser wie z.B. Internet Explorer, Chrome, Opera oder Safari war weniger erfolgreich. Sobald die Anforderungen an einen Browser in Sachen Layout, Design oder Berechnungen per JavaScript aufwendiger und komplexer werden, zeigt sich auch die Problematik mit der browserübergreifenden Entwicklung. Die meisten Probleme in der Webentwicklung liefert meist der Internet Explorer. Daher werden in weiterer Folge auch speziell die in dem Browser aufgetretenen Probleme beschrieben.

5.1.10.1 Probleme im IE (Internet Explorer)

Da der online EPUB Reader Divs verwendet, die selbst noch mit CSS Eigenschaften belegt sind, hatte der Internet Explorer Probleme diese darzustellen. Es kam des öfteren vor, dass der IE in den »Quirks - Modus« ging, um den online EPUB Reader anzuzeigen. Dieser Modus dient dazu, die Abwärtskompatibilität veralteter Befehle oder Eigenschaften sicher zu stellen. Grund dafür war eine fehlende »doctype« Angabe im Code. Durch diese Änderung wurde ein Teil des Layouts korrekt wiedergegeben. Da die für die Berechnung der Seitenanzahl fixe Höhe und Breite der Divs vorgegeben sind, gab es auch hier Darstellungsprobleme.

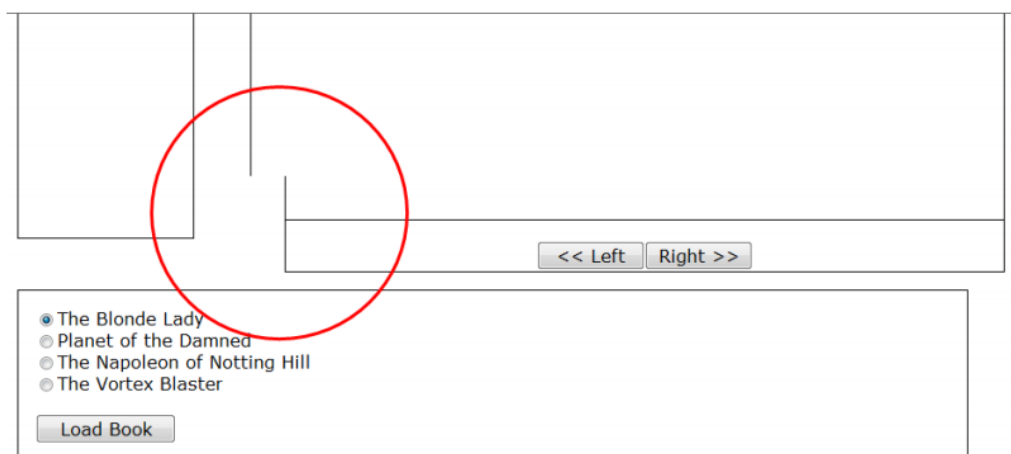


Abbildung 5.13: Fehlerhafte Darstellung der Divs im Internet Explorer

Um diesen Fehler zu beheben, wurde in der CSS Definition eine Browserweiche mit eigenen Befehlen definiert. Dazu wird zuerst der Typ des Browser abgefragt. Trifft der Browsertyp zu, wird anstatt der globalen CSS Definition eine speziell für diesen Browser definierte CSS Definition aufgerufen.

```
<!--[if IE]>
  <style type="text/css" >

      #wrapper {
        width: 882px;
      }
  </style>
</if IE-->
```

Quellcode 5.5: Speziell angepasstes CSS für Internet Explorer

Aufgrund der für den IE angepassten Höhen und Breiten musste auch der Code zur Seitenzahlberechnung speziell für diesen Browser angepasst werden. Der globale Code liefert sonst fehlerhafte Ergebnisse.

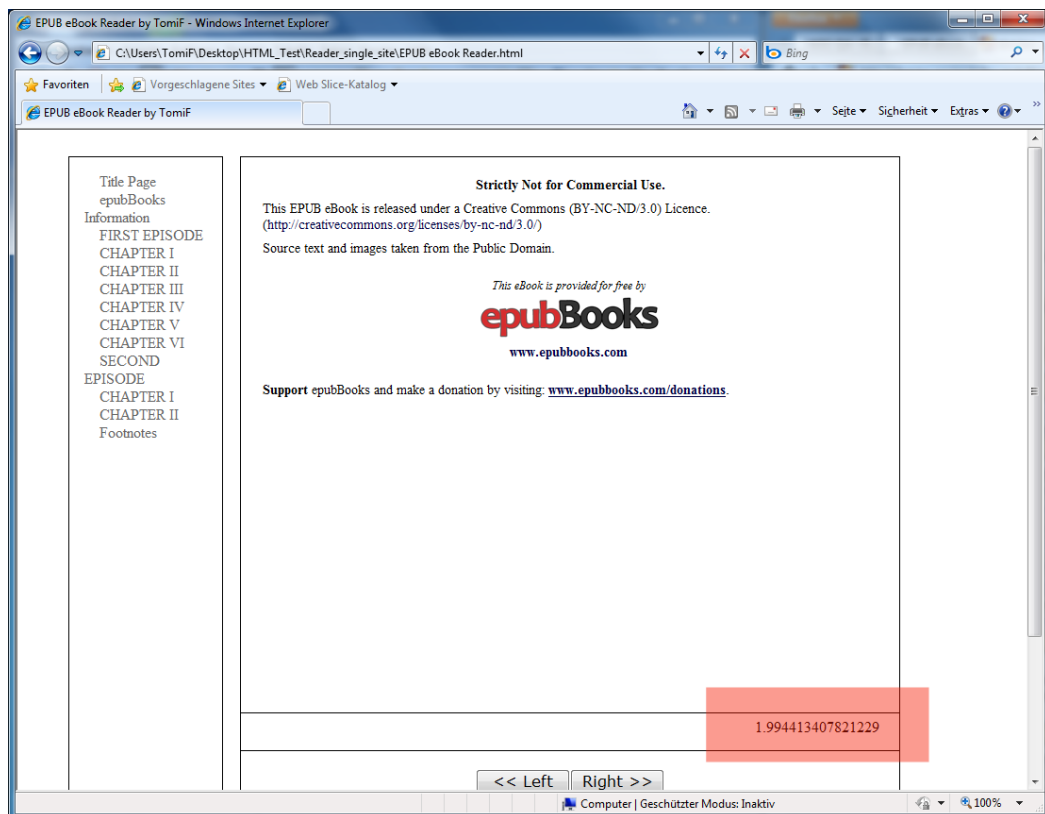


Abbildung 5.14: Fehlerhafte Seitenberechnung im Internet Explorer

5.1.10.2 Fehlende CSS3 Unterstützung

Ein weiteres Problem war die unzureichende oder unterschiedliche Implementierung von CSS3 Befehlen. Gecko und Webkit basierte Browser bieten eigene Interpretationen von CSS3 Befehlen an. Diese sind durch Präfixe wie »-moz« oder »-webkit« gekennzeichnet und aufrufbar. Leider war diese Unterstützung beim Internet Explorer zum Zeitpunkt der Entwicklung noch nicht gegeben.

Aufgrund dieser Einschränkung wurde versucht, die spaltenweise Formatierung im Internet Explorer auf andere Weise zu implementieren. Es wurden dazu verschiedene jQuery Plugins kombiniert und getestet.

- jQuery Plugin Columnize [[jQu11](#)]
- jQuery Plugin Columnizer [[Wulo8](#)]
- jQuery CSS3 Multi-Column Module [[Savo5](#)]

Einiger dieser Plugin lieferten zwar gute Ergebnisse, reichen aber nicht an die natürliche Implementierung des CSS3 Befehls »multi-column« heran. Durch die umfangreiche Vorabberechnung der Seiten und Strukturierung sind einige Plugins nicht mehr performant genug, um im online EPUB Reader verwendet zu werden.

Aufgrund dieser Einschränkung ist der volle Funktionsumfang des online EPUB Readers erst im Internet Explorer erst ab der Version 10 gegeben. [[Mic12](#)]

6 Zusammenfassung und Ausblick

Teil dieser Arbeit war es, Technologien, Konzepte und Formate aufzuzeigen, die in Verbindung mit digitalen Bibliotheken entwickelt wurden. Ein weiterer Punkt war die Implementierung eines Readers, mit dem es möglich ist, das für mobile Geräte entwickelte Format EPUB, ohne zusätzliche Hilfsmittel im Browser darzustellen.

Es wird gezeigt, dass durch die technologische Entwicklung einige Probleme klassischer Bibliotheken gelöst werden konnten, gleichzeitig aber neue Probleme und Fragen entstanden sind. Vor allem das Problem der Langzeitarchivierung digitaler Medien wird ausführlich diskutiert und mit Beispielen unterlegt.

Des Weiteren werden gängige Anzeigeformate untersucht, deren Vor- und Nachteile aufgezeigt und Eigenschaften erörtert, die notwendig sind, um in digitalen Bibliotheken verwendet zu werden. Es werden Möglichkeiten aufgezeigt, wie DRM im Umfeld von digitalen Bibliotheken eingesetzt werden kann. In Bezug auf DRM wird dessen fragwürdiger Einsatz in ähnlichen Bereichen wie z.B. der Musikindustrie herangezogen und kritisch hinterfragt. Darauf folgend werden verschiedene Metadaten und Austauschprotokolle beschrieben, die vor allem im Bereich digitaler Bibliotheken und Archiven ihren Einsatz finden.

Anschließend wird OpenSource-Software bewertet, mit der es möglich ist, Eigenschaften digitaler Bibliotheken wie z.B. Erfassung, Speicherung und Weiterverbreitung digitaler Ressourcen durchzuführen. Die Bewertung erfolgt durch vorab festgelegte Fragen, die bei einer möglichen Einführung auftreten würden.

Abschließend erfolgt die Implementierung eines online EPUB Readers, mit dem es möglich ist, EPUB Formate im Browser darzustellen. Es wird gezeigt, dass ein für mobile Geräte entwickeltes Format auch auf Desktopgeräten angezeigt werden kann. Voraussetzung dafür ist, dass das Format offen und deren Aufbau gut dokumentiert ist. Es ist dadurch möglich, auch ohne mobiles Gerät und ohne zusätzliche Installationen oder Browser-Plugins in einer EPUB Publikation zu lesen.

Die Formate im Umfeld digitaler Bibliotheken zeigen deutlich in Richtung Mobilität. In den letzten Jahren war vor allem in den Bereichen Smartphones, Tablets

und EReader ein besonderer Technologie- und Entwicklungsschub erkennbar. Dies war Voraussetzung für neue Formate und Hardware, die das Lesen von digitalen Medien revolutionierten. Diese Innovation ist derzeit leider nur auf mobile Endgeräte beschränkt. Viele Verlage und Bibliotheken erkennen erst jetzt das Potential dieser Entwicklung und beginnen nun sukzessive mit dem Ausbau von Serviceangeboten.

Appendix

Literatur

- [Adoo8] Adobe. *Document management — Portable document format — Part 1: PDF 1.7*. aufgerufen am 08.05.2012. 2008. URL: http://www.adobe.com/content/dam/Adobe/en/devnet/acrobat/pdfs/PDF32000_2008.pdf.
- [AFo5] Caroline Arms und Carl Fleischhauer. *Digital Formats: Factors for Sustainability, Functionality, and Quality*. Report. Office of Strategic Initiatives, Library of Congress Washington, DC, USA, 2005. URL: http://memory.loc.gov/ammem/techdocs/digform/Formats_IST05_paper.pdf.
- [Ama12a] Amazon. *Customer Reviews - Spore*. aufgerufen am 17.05.2012. 2012. URL: http://www.amazon.com/Spore-Mac/product-reviews/B000FKBCX4/ref=sr_1_1_cm_cr_acr_txt?ie=UTF8&showViewpoints=1.
- [Ama12b] Amazon. *Kindle Format 8*. aufgerufen am 10.05.2012. 2012. URL: <http://www.amazon.com/gp/feature.html?docId=1000729511>.
- [Arm99] William Y. Arms. *Digital Libraries (Digital Libraries and Electronic Publishing)*. aufgerufen am 22.05.2012. The MIT Press, 1999. ISBN: 0262011808. URL: <http://www.cs.cornell.edu/wya/DigLib/MS1999/index.html>.
- [Bal11] Matthias Ballod. *Informationen und Wissen im Griff : effektiv informieren und effizient kommunizieren*. Bielefeld: Bertelsmann, 2011. ISBN: 9783763936977.
- [BBC12] BBC. *Story of Domesday*. aufgerufen am 03.04.2012. 2012. URL: <http://www.bbc.co.uk/history/domesday/story>.
- [Ber11] René Berndt. *Digitale Bibliotheken - Aufbau einer DL*. aufgerufen am 19.07.2012. 2011.
- [Bib11] Deutscher Bibliotheksverband. *Bericht zur Lage der Bibliotheken 2011. Zahlen und Fakten*. Deutscher Bibliotheksverband e.V. (dbv), 2011. URL: http://www.bibliotheksverband.de/fileadmin/user_upload/DBV/themen/Bericht_zur_Lage_der_Bibliotheken_2011_Web.pdf.

- [Blo+05] Barbara Block u. a. *Maschinelles Austauschformat für Bibliotheken und die Functional Requirements for Bibliographic Records*. Report. Berlin-Dahlem Germany: Konrad-Zuse-Zentrum für Informationstechnik Berlin, 2005. URL: <http://www.zib.de/Publications/Reports/ZR-05-60.pdf>.
- [Can+07] L. Candela u. a. *The DELOS Digital Library Reference Model. Foundations for Digital Libraries*. DELOS Network of Excellence on Digital Libraries, 2007, S. 17. URL: http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf.
- [CEN04] CEN/ISSS. *Europäischer Leitfaden zur erfolgreichen Praxis im Wissensmanagement*. Workshop. Brüssel: CEN/ISSS, 2004. Kap. 1, S. 10. URL: <ftp://cenftp1.cenorm.be/PUBLIC/CWAs/e-Europe/KM/German-text-KM-CWAGuide.pdf>.
- [Chu+08] Christopher Chute u. a. *The Diverse & Exploding Digital Universe. An Updated Forecast of Worldwide Information Growth Through 2011*. Techn. Ber. IDC, 2008, S. 16. URL: <http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf>.
- [Con11] Library of Congress. Z39.50. aufgerufen am 31.05.2012. 2011. URL: <http://www.loc.gov/z3950/agency/>.
- [Con12a] Library of Congress. *MARC - Standards*. aufgerufen am 23.05.2012. 2012. URL: <http://www.loc.gov/marc/>.
- [Con12b] Library of Congress. *MARCXML - Marc 21 XML Schema*. aufgerufen am 23.05.2012. 2012. URL: <http://www.loc.gov/standards/marcxml/>.
- [Con12c] Library of Congress. *Metadata Encoding & Transmission Standard*. aufgerufen am 23.05.2012. 2012. URL: <http://www.loc.gov/standards/mets/>.
- [Con12d] Library of Congress. *MODS - Metadata Object Description Schema*. aufgerufen am 23.05.2012. 2012. URL: <http://www.loc.gov/standards/mods/>.
- [Con12e] Library of Congress. *Search/Retrieval via URL*. aufgerufen am 31.05.2012. 2012. URL: <http://www.loc.gov/standards/sru/>.
- [Deg12] Andreas Degkwitz. *Die digitale Revolution erfasst die Bibliotheken*. aufgerufen am 06.03.2012. 2012. URL: <http://www.tagesspiegel.de/wissen/bibliotheken-die-digitale-revolution-erfasst-die-bibliotheken/6010280.html>.

- [der12] derStandard.at. *Forscher boykottieren Fachverlag Elsevier*. aufgerufen am 11.04.2012. 2012. URL: <http://derstandard.at/1329703169887/Teure-Wissenschaft-Forscher-boykottieren-Fachverlag-Elsevier>.
- [Dig09] Mark Diggory. *DSpace System Documentation*. aufgerufen am 16.07.2012. 2009. URL: https://wiki.duraspace.org/download/attachments/22022817/DSpace_System_Documentation_1.6.0_rc1.pdf.
- [Dlu12] Vladislav Dluzewski. *Fragen zur Datenrettung*. aufgerufen am 17.04.2012. 2012. URL: <http://www.hddl.de/datenrettung/datenrettung.html>.
- [Dob07] Susanne Dobratz. *Open-Source-Software zur Realisierung von Institutionellen Repositories – Überblick*. Techn. Ber. Humboldt-Universität zu Berlin, 2007, S. 205.
- [dub10] dublicore.org. *Dublin Core Metadata Element Set, Version 1.1*. aufgerufen am 23.05.2012. 2010. URL: <http://dublicore.org/documents/dces/>.
- [Dur12] Duraspace. *Duraspace - About*. aufgerufen am 04.06.2012. 2012. URL: <http://duraspace.org/about.php>.
- [DW11] Daniel Davis und Chris Wilper. *Fedora Digital Object Model*. aufgerufen am 26.07.2012. 2011. URL: <https://wiki.duraspace.org/display/FEDORA34/Fedora+Digital+Object+Model>.
- [Efoo] Albert Endres und Dieter W Fellner. *Digitale Bibliotheken : Informatik Lösungen für globale Wissensmärkte*. Heidelberg: Dpunkt, 2000. ISBN: 3932588770 9783932588778.
- [Fed12] Fedora Commons. *Fedora Commons Repository Software - open source technologies to manage, preserve and link your digital content*. aufgerufen am 04.06.2012. 2012. URL: <http://www.fedora-commons.org/>.
- [For98] Deutsche Forschungsgemeinschaft. *Vorschläge zur Sicherung guter wissenschaftlicher Praxis*. WILEY-VCH Verlag GmbH, 1998. ISBN: 3-527-27212-7. URL: http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf#page=17.
- [Fou12a] DSpace Foundation. *DSpace*. aufgerufen am 16.07.2012. 2012. URL: <http://www.dspace.org/>.
- [Fou12b] jQuery Foundation. *jQuery is a new kind of JavaScript Library*. aufgerufen am 21.06.2012. 2012. URL: <http://jquery.com/>.
- [Fou12c] The Document Foundation. *LibreOffice*. aufgerufen am 29.03.2012. 2012. URL: <http://de.libreoffice.org/>.

- [Fun12a] Stefan E. Funk. »Emulation«. In: *nestor Handbuch*. Hrsg. von H. Neuroth u. a. Universität Göttingen, 2012. Kap. 8.16. URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_403.pdf.
- [Fun12b] Stefan E. Funk. »Migration«. In: *nestor Handbuch*. Hrsg. von H. Neuroth u. a. 2012. Kap. 8.10. URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_450.pdf.
- [Gab12] Gabler. *Digital Rights Management*. aufgerufen am 16.05.2012. 2012. URL: <http://wirtschaftslexikon.gabler.de/Archiv/75892/digital-rights-management-v7.html>.
- [Gam12] Gameshouse. *Weltweit erstes Computerspielemuseum*. aufgerufen am 29.03.2012. 2012. URL: http://www.computerspielemuseum.de/1219_Ueber_uns.htm.
- [Gla12] Glazou. *iBooks Author, a nice tool but..* aufgerufen am 17.05.2012. 2012. URL: <http://www.glazman.org/weblog/dotclear/index.php?post/2012/01/20/iBooks-Author-a-nice-tool-but>.
- [Gol12] Golem. *Apple sabotiert ePub-Format mit iBooks Author*. aufgerufen am 10.05.2012. 2012. URL: <http://www.golem.de/1201/89261.html>.
- [Gut12] Gutenberg. *Project Gutenberg - the first producer of free ebooks*. aufgerufen am 23.04.2012. 2012. URL: <http://www.gutenberg.org/wiki/Gutenberg>About>.
- [Heio3] Heise. *Berichte über DRM-Probleme bewirken Rückzieher von Apple*. aufgerufen am 16.05.2003. 2003. URL: <http://www.heise.de/newsticker/meldung/Berichte-ueber-DRM-Probleme-bewirken-Rueckzieher-von-Apple-82787.html>.
- [Heuo3] Sylke Heun. *Fünf Millionen Berliner Bücher durch Säurefraß bedroht*. aufgerufen am 19.03.2012. 2003. URL: http://www.welt.de/print-welt/article249242/Fuenf_Millionen_Berliner_Buecher_durch_Saeurefrass_bedroht.html.
- [Hir96] Michael Christian Hirsch. »Subito - ein Schlagwort und was dahinter steht«. In: *Die Herausforderungen der Bibliotheken durch elektronische Medien und neue Organisationsformen 85*. Deutscher Bibliothekartag. Frankfurt am Main: Wefers, Sabine, 1996, S. 67 –72.
- [HL07] Martin Hilbert und Priscila Lopez. *The world's technological capacity to process information*. aufgerufen am 01.03.2012. 2007. URL: <http://www.martinhilbert.net/WorldInfoCapacity.html>.

- [Hof12a] Nate Hoffelder. *Amazon Saw a 5-Fold Increase in UK Ebook Sales*. aufgerufen am 24.04.2012. 2012. URL: <http://www.the-digital-reader.com/2012/02/01/amazon-saw-a-5-fold-increase-in-uk-ebook-sales/>.
- [Hof12b] Andreas Hofmann. *Migrationslose Archivierung digitaler Daten*. aufgerufen am 05.04.2012. Frauenhofer. 2012. URL: http://www.ipm.fraunhofer.de/de/Loesungen_Maerkte/Medien_Kommunikation/sichere_langzeitarchivierung/bits_on_film.html.
- [Hur+99] Bernard J. Hurley u. a. *The Making of America II Testbed Project*. aufgerufen am 22.05.2012. Washington, D.C.: Digital Library Federation, 1999. ISBN: 978-1-933645-04-9. URL: <http://old.diglib.org/pubs/dlf089/dlf089.pdf>.
- [Hut12] Karsten Hut. »Computermuseum«. In: *nestor Handbuch*. Hrsg. von H. Neuroth u. a. Universität Göttingen, 2012. Kap. 8.24. URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_451.pdf.
- [IDP12a] IDPF. *EPUB 3*. aufgerufen am 02.05.2012. 2012. URL: <http://idpf.org/epub/30>.
- [IDP12b] IDPF. *EPUB Content Documents 3.0*. aufgerufen am 02.05.2012. 2012. URL: <http://idpf.org/epub/30/spec/epub30-contentdocs.html>.
- [IDP12c] IDPF. *EPUB Media Overlays 3.0*. aufgerufen am 02.05.2012. 2012. URL: <http://idpf.org/epub/30/spec/epub30-mediaoverlays.html>.
- [IDP12d] IDPF. *EPUB Open Container Format (OCF) 3.0*. aufgerufen am 02.05.2012. 2012. URL: <http://idpf.org/epub/30/spec/epub30-ocf.html>.
- [IDP12e] IDPF. *EPUB Publications 3.0*. aufgerufen am 02.05.2012. 2012. URL: <http://idpf.org/epub/30/spec/epub30-publications.html>.
- [IDP12f] IDPF. *Member List*. aufgerufen am 03.05.2012. 2012. URL: <http://idpf.org/membership/members>.
- [ISOo6] ISO/IEC. *Information technology - Open Document Format for Office Applications (OpenDocument) v1.0*. aufgerufen am 20.04.2012. 2006. URL: <http://standards.iso.org/ittf/licence.html>.
- [ISOo8] ISO/IEC. *Information technology – Document description and processing languages – Office Open XML File Formats*. aufgerufen am 20.04.2012. 2008. URL: http://www.iso.org/iso/catalogue_detail?csnumber=51463.
- [jQu11] jQuery. *jQuery Plugins*. aufgerufen am 28.07.2011. 2011. URL: <http://plugins.jquery.com/>.

- [JS12] Mathias Jehn und Sabine Schrimpf. »LZA-Aktivitäten in Deutschland aus dem Blickwinkel von nedor«. In: *nedor Handbuch*. Hrsg. von H. Neuroth u. a. Universität Göttingen, 2012. Kap. 2.2. URL: http://nedor.sub.uni-goettingen.de/handbuch/artikel/nedor_handbuch_artikel_391.pdf.
- [Kei12] Christian Keitel. »Mikroverfilmung«. In: *nedor Handbuch*. Hrsg. von H. Neuroth u. a. Universität Göttingen, 2012. Kap. 8.32. URL: http://nedor.sub.uni-goettingen.de/handbuch/artikel/nedor_handbuch_artikel_452.pdf.
- [Kök11] Michal Kökörčény. »Architecture and flexibility of digital libraries systems«. In: *INTERNATIONAL JOURNAL OF SYSTEMS APPLICATIONS, ENGINEERING u. DEVELOPMENT* 5 (2 2011).
- [Köl12] Bernhard Kölbl. *Der Kulturbote lebt - und engagiert sich*. aufgerufen am 04.04.2012. 2012. URL: <http://derkulturbote.de/>.
- [Lag+02] Carl Lagoze u. a. *The Open Archives Initiative Protocol for Metadata Harvesting*. aufgerufen am 31.05.2012. 2002. URL: <http://www.openarchives.org/OAI/openarchivesprotocol.html>.
- [Les97] Michael Lesk. *Practical Digital Libraries*. Morgan Kaufmann Publishers, Inc., 1997. Kap. 9, S. 215. ISBN: 1558604596.
- [Leu09] Laurens Leurs. *An introduction to PDF*. aufgerufen am 26.04.2012. 2009. URL: <http://www.prepressure.com/pdf/basics/introduction>.
- [LK08] Thomas Lennartz und Gordon Krause. *Deutsche Fachpresse - Digital Rights Management*. aufgerufen am 16.05.2012. 2008. URL: http://www.deutsche-fachpresse.de/fileadmin/allgemein/downloads/DRM-Praesentation_2008_final.ppt.
- [Mad+09] Olivia Madison u. a. *Functional Requirements for Bibliographic Records*. Final Report. International Federation of Library Associations und Institutions, 2009. URL: http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf.
- [Maz09] Riccardo Mazza. *Introduction to information visualization*. Springer, 2009.
- [Melo3] Phil Mellor. *CAMiLEON: Emulation and BBC Domesday*. aufgerufen am 03.04.2012. 2003. URL: <http://www.iconbar.com/forums/viewthread.php?newsid=937>.
- [Mic10] Microsoft. *Microsoft Office Binary (doc, xls, ppt) File Formats*. aufgerufen am 20.04.2012. 2010. URL: <http://www.microsoft.com/interop/docs/OfficeBinaryFormats.msp>.

- [Mic12] Microsoft. *Internet Explorer 10 Guide for Developers - Multi-column layout*. aufgerufen am 05.07.2012. 2012. URL: [http://msdn.microsoft.com/en-us/library/ie/hh673534\(v=vs.85\).aspx](http://msdn.microsoft.com/en-us/library/ie/hh673534(v=vs.85).aspx).
- [Mob12] MobiPocket. *Mobipocket.com*. aufgerufen am 10.05.2012. 2012. URL: <http://www.mobipocket.com/en/Corporate/AboutMobipocket.asp?Language=DE>.
- [Moz12] Mozilla. *Firebug - Web Development Evolved*. aufgerufen am 04.07.2012. 2012. URL: <http://getfirebug.com/whatisfirebug>.
- [MS97] William J. Mitchell und Oliver Strimpel. »There and not there«. In: *Beyond calculation*. Hrsg. von Peter J. Denning und Robert M. Metcalfe. New York, NY, USA: Copernicus, 1997, S. 245–258. ISBN: 0-38794932-1. URL: <http://dl.acm.org/citation.cfm?id=504928.504946>.
- [MW11] Uwe Matrisch und Ursula Welsch. *ePUB 3 – Eine kleine Übersicht*. aufgerufen am 08.05.2012. 2011. URL: <http://e-book-corner.blogspot.com/2011/07/epub-3-eine-kleine-ubersicht.html>.
- [Nat12a] Österreichische Nationalbibliothek. *Ablieferung digitaler Medien*. aufgerufen am 13.03.2012. 2012. URL: http://www.onb.ac.at/bibliothek/ablieferung_digitaler_medien.htm.
- [Nat12b] Österreichische Nationalbibliothek. *Pflichtablieferung – Legal deposit – Dépôt légal*. aufgerufen am 13.03.2012. 2012. URL: <http://www.onb.ac.at/bibliothek/pflichtablieferung.htm>.
- [NISO4] NISO. *Understanding metadata*. aufgerufen am 22.05.2012. National Information Standards Organization. 2004. ISBN: 1-880124-62-9. URL: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>.
- [noo12] nooxml.org. *Say NN to Microsoft Office broken standard*. aufgerufen am 20.04.2012. 2012. URL: <http://www.nooxml.org/rice-pudding>.
- [OAI12] OAI. *Open Archives Initiative*. aufgerufen am 31.05.2012. 2012. URL: <http://www.openarchives.org/>.
- [OPD12a] OPDS. *OPDS Catalog 1.1*. aufgerufen am 12.09.2012. 2012. URL: <http://opds-spec.org/specs/opds-catalog-1-1-20110627/>.
- [OPD12b] OPDS. *Open Publication Distribution System*. aufgerufen am 12.09.2012. 2012. URL: <http://opds-spec.org/>.
- [ORe12a] Jay O’Rear. *How DRM works*. aufgerufen am 16.05.2012. 2012. URL: <http://www.adobe.com/uk/epaper/features/drm/howdrmworks.html>.

- [ORe12b] O'Reilly. *Ebook Usage, Devices, and Formats*. aufgerufen am 16.05.2012. 2012. URL: <http://shop.oreilly.com/category/customer-service/ebooks.do>.
- [Pan10] Christian Pansch. *Vorteile von DRM für den Rechteinhaber/User*. aufgerufen am 17.05.2012. 2010. URL: <http://www.christian-pansch.de/mein-wissen/digital-rights-management-durchleuchtet/vorteile-von-drm/>.
- [Ped12] Chris Pederick. *Web Developer*. aufgerufen am 04.07.2012. 2012. URL: <http://chrispederick.com/work/web-developer/>.
- [PN07] George Pirounakis und Mara Nikolaidou. »Comparing Open Source Digital Library Software«. In: (2007). URL: <http://www.dit.hua.gr/~mara/publications/ideaDL09a.pdf>.
- [Pra12] Mandi Pralle. *The Advantages and Disadvantages of Using the EPUB Format*. aufgerufen am 03.05.2012. 2012. URL: <http://articles.submyourarticle.com/Mandi-Pralle-6459/epub-document-format-207622.php>.
- [Ran07] Abby Randal. *Why To Convert Word To PDF - 10 Advantages Of PDF Files*. aufgerufen am 26.04.2012. 2007. URL: <http://ezinearticles.com/?Why-To-Convert-Word-To-PDF---10-Advantages-Of-PDF-Files&id=485574>.
- [Ran31] Shiyali Ramamrita Ranganathan. *The five laws of library science*. University of California: The Madras Library Association, 1931, S. 458.
- [Rap11] Lisa Rapaport. *Amazon.com Says Kindle E-Book Sales Surpass Printed Books for First Time*. aufgerufen am 24.04.2012. 2011. URL: <http://www.bloomberg.com/news/2011-05-19/amazon-com-says-kindle-electronic-book-sales-surpass-printed-format.html>.
- [RB10] Jenn Riley und Devin Becker. *Seeing Standards: A Visualization of the Metadata Universe*. aufgerufen am 29.05.2012. Indiana University. 2010. URL: <http://www.dlib.indiana.edu/~jenlrile/metadatamap/>.
- [RW07] Judith Rog und Caroline van Wijk. *Evaluating File Formats for Long-term Preservation*. Report. National Library of the Netherlands, 2007. URL: http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/KB_file_format_evaluation_method_27022008.pdf.
- [Sai12] Saikat. *What is the Difference Between EPUB, MOBI, AZW and PDF eBook Formats?* aufgerufen am 10.05.2012. 2012. URL: <http://www.guidingtech.com/9661/difference-between-epub-mobi-azw-pdf-ebook-formats/>.

- [Savo5] Cedric Savarese. *Introducing the CSS3 Multi-Column Module*. aufgerufen am 05.07.2012. 2005. URL: <http://www.alistapart.com/articles/css3multicolumn>.
- [Scho5] Christopher Schrader. *Aus dem Regal auf die Festplatte*. aufgerufen am 17.04.2012. 2005. URL: <http://www.sueddeutsche.de/digital/digitale-buecher-aus-dem-regal-auf-die-festplatte-1.835269>.
- [Sch10] Thomas Schwenke. *Buch der Zukunft: Über die leise und unaufhaltsame Revolution im Verlagswesen*. Hrsg. von Jan Tißler. Tredition, Sep. 2010, S. 201. ISBN: 9783868508307. URL: <http://amazon.de/o/ASIN/3868508309/>.
- [Sch12] Frederick Schiwek. *Gelöschte Dateien wiederherstellen*. aufgerufen am 17.04.2012. 2012. URL: <http://datenverlust.org/geloeschte-dateien-wiederherstellen/>.
- [Sfe11] Nicolae Sfetcu. *eBook timeline*. aufgerufen am 24.04.2012. 2011. URL: <http://www.bestdownload.biz/ebooks/tag/digital-book-format/>.
- [Sha12] Pallavi Shah. *Digital Rights Management - What is?, Why DRM?, Technology/Architecture Overview, Selection Guidelines, DRM vs. Contracts Rights Management*. aufgerufen am 16.05.2012. 2012. URL: <http://encyclopedia.jrank.org/articles/pages/6724/Digital-Rights-Management.html>.
- [She99] Nathan Shedroff. »Information interaction design: A unified field theory of design«. In: *Information Design*. Hrsg. von Robert Jacobson. MIT Press, 1999. Kap. 11, S. 267–292.
- [Soco5] The Internet Society. *The Atom Syndication Format*. aufgerufen am 17.09.2012. 2005. URL: <http://www.ietf.org/rfc/rfc4287.txt>.
- [Sto09] Brad Stone. *Amazon Erases Orwell Books From Kindle*. aufgerufen am 17.05.2012. 2009. URL: http://www.nytimes.com/2009/07/18/technology/companies/18amazon.html?_r=3.
- [SWD09] Martin Steinebach, Patrick Wolf und Konstantin Diener. *Urheber-schutz im Internet: CoSee findet wasserzeichenmarkierte Medien - Medien aufspüren, Wasserzeichen auslesen, Piraten verfolgen - Kundenfreundliche Alternative zu teurer DRM-Technik*. aufgerufen am 16.05.2012. 2009. URL: http://www.mpr-frankfurt.de/presse/sit/090625_fraunhofer_sit_cosee_pressemitteilung.htm.

- [Tho11] Ted Thornhill. *BBC Domesday Project re-released on touchscreen computer after laser disc archive fails*. aufgerufen am 03.04.2012. BBC. 2011. URL: <http://www.dailymail.co.uk/sciencetech/article-2072073/BBC-Domesday-Project-released-touchscreen-laser-disc-archive-fails.html>.
- [Ull12] Dagmar Ullrich. »Bitstream Preservation«. In: *nestor Handbuch*. Hrsg. von H. Neuroth u. a. Universität Göttingen, 2012. Kap. 8.3. URL: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_402.pdf.
- [Vee+12] Peter Veenstra u. a. *DOSBox*. aufgerufen am 29.03.2012. 2012. URL: <http://www.dosbox.com/information.php?page=0>.
- [W3C11] W3C. *CSS Multi-column Layout Module*. aufgerufen am 27.06.2012. 2011. URL: <http://www.w3.org/TR/css3-multicol/>.
- [Walo8] Chris Walters. *Amazon Pulls Negative Reviews Of 'Spore,' Then Reinstates Them*. aufgerufen am 17.05.2012. 2008. URL: <http://consumerist.com/2008/09/amazon-pulls-negative-reviews-of-spore-then-reinstates-them.html>.
- [Wik12a] Wikipedia. *BBC Domesday Project*. aufgerufen am 03.04.2012. 2012. URL: http://en.wikipedia.org/wiki/BBC_Domesday_Project.
- [Wik12b] Wikipedia. *Daten*. aufgerufen am 06 03 2012. 2012. URL: <http://de.wikipedia.org/wiki/Daten>.
- [Wik12c] Wikipedia. *Digitale Rechteverwaltung*. aufgerufen am 17.05.2012. 2012. URL: http://de.wikipedia.org/wiki/Digitale_Rechteverwaltung.
- [Wik12d] Wikipedia. *Domesday Book*. aufgerufen am 03.04.2012. 2012. URL: http://de.wikipedia.org/wiki/Domesday_Book.
- [Wik12e] Wikipedia. *Elektronisches Papier*. aufgerufen am 11 04 2012. 2012. URL: http://de.wikipedia.org/wiki/Elektronisches_Papier.
- [Wik12f] Wikipedia. *Information*. aufgerufen am 06 03 2012. 2012. URL: <http://de.wikipedia.org/wiki/Information>.
- [Wik12g] Wikipedia. *iTunes*. aufgerufen am 2012. 2012. URL: http://de.wikipedia.org/wiki/iTunes#cite_note-12.
- [Wik12h] Wikipedia. *iTunes Store*. aufgerufen am 16.05.2012. 2012. URL: http://de.wikipedia.org/wiki/iTunes_Store.
- [Wik12i] Wikipedia. *Metadaten*. aufgerufen am 18.05.2012. 2012. URL: <http://de.wikipedia.org/wiki/Metadaten>.
- [Wik12j] Wikipedia. *Publish or perish*. aufgerufen am 22 03 2012. 2012. URL: http://de.wikipedia.org/wiki/Publish_or_perish.

- [Wulo8] Adam Wulf. *Columnizer jQuery Plugin*. aufgerufen am 05.07.2012. 2008. URL: <http://welcome.totheinter.net/columnizer-jquery-plugin/>.