

**Web-Site-Biographie – Anforderungsanalyse,  
Systementwurf, prototypische Realisierung**

**Dissertation**

am Fachbereich Wirtschaftswissenschaften  
an der Justus-Liebig-Universität Gießen

zur Erlangung des Grades eines  
Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)

vorgelegt von:

Maximilian Patzak, M. Sc.

aus Gießen

20. November 2020

# Inhaltsverzeichnis

	Seite
Abbildungsverzeichnis .....	IV
Abkürzungsverzeichnis .....	VI
<b>1 Problemstellung, Ziel und Aufbau der Arbeit.....</b>	<b>1</b>
1.1 Problemstellung und Motivation der Arbeit .....	1
1.2 Untersuchungsbereich und Zielsetzung .....	10
1.3 Forschungsmethode und Aufbau der Arbeit .....	11
<b>2 Definition, Abgrenzung, Nutzen, Umsetzung einer WSB.....</b>	<b>15</b>
2.1 Systematisierung der Grundlagen.....	15
2.2 Definition einer „Web-Site-Biographie“ .....	15
2.3 Abgrenzung einer Web-Site-Biographie.....	20
2.4 Zum Nutzen einer Web-Site-Biographie .....	22
2.5 Aufgaben zur systembasierten Umsetzung einer WSB .....	26
2.5.1 Systematisierung.....	26
2.5.2 Erfassen.....	28
2.5.3 Bewahren .....	32
2.5.4 Bereitstellen.....	36
<b>3 Eigenschaftsanalyse einer Web Site .....</b>	<b>38</b>
3.1 Systematisierung der Eigenschaftsanalyse.....	38
3.2 Inhaltliche Eigenschaften.....	39
3.2.1 Flüchtigkeit .....	39
3.2.2 Interaktivität und Absicherung.....	40
3.2.3 Kritikalität und Personalisierung.....	43
3.3 Gestalterische Eigenschaften.....	46
3.3.1 Grafische Aufbereitung und Positionierung .....	46
3.3.2 Web-Site-Strukturierung und Vielgestaltigkeit.....	48
3.4 Betriebsspezifische Eigenschaften .....	50
3.4.1 Formatvielfalt.....	50
3.4.2 Infrastrukturintegration und Infrastrukturwandel.....	51
3.4.3 Verteilung, Sicherheit und Interoperabilität.....	53
3.5 Zusammenfassung und Erläuterung des Fortgangs der Arbeit .....	55

<b>4</b>	<b>Anforderungsanalyse eines WSBS</b> .....	<b>56</b>
4.1	Systematisierung der Anforderungsanalyse .....	56
4.2	Fachliche Anforderungsanalyse .....	57
4.2.1	Systematisierung der fachlichen Anforderungsanalyse .....	57
4.2.2	Anforderungen auf Basis inhaltlicher Eigenschaften .....	58
4.2.3	Anforderungen auf Basis gestalterischer Eigenschaften .....	62
4.2.4	Anforderungen auf Basis betriebsspezifischer Eigenschaften .....	65
4.3	Organisatorische Anforderungsanalyse .....	69
4.3.1	Systematisierung der organisatorischen Anforderungsanalyse .....	69
4.3.2	Strukturorientierte Anforderungen .....	69
4.3.3	Prozessorientierte Anforderungen .....	70
4.4	Technische Anforderungsanalyse .....	73
4.4.1	Systematisierung der technischen Anforderungsanalyse .....	73
4.4.2	Standardisierte Qualitätsanforderungen .....	74
4.4.3	Anforderungen an den Erhalt digitaler Abbilder .....	77
4.5	Zusammenfassung und Erläuterung des Fortgangs der Arbeit .....	79
<b>5</b>	<b>Systementwurf eines WSBS</b> .....	<b>81</b>
5.1	Systematisierung des Systementwurfs .....	81
5.2	Systemarchitektur: Systembausteine eines WSBS .....	83
5.2.1	Systematisierung der Systemarchitektur .....	83
5.2.2	Schichtenarchitektur und Systembausteine .....	85
5.2.3	Datenhaltungsschicht .....	87
5.2.4	Anwendungen der Steuerungsschicht .....	89
5.2.5	Bedienoberflächen der Präsentationsschicht .....	93
5.2.6	Verfügbarkeitsdienste .....	98
5.3	Determinanten der Datenmodellierung: Abbilder und Metadaten .....	100
5.3.1	Systematisierung der Datenmodellierung .....	100
5.3.2	Abbilder .....	101
5.3.3	Metadaten .....	106
5.3.4	Zur Implementierung eines Datenmodells .....	109
5.4	Module und Funktionen .....	111
5.4.1	Systematisierung der Module und Funktionen .....	111
5.4.2	Erfassen .....	112
5.4.3	Bewahren .....	116
5.4.4	Bereitstellen .....	121
5.4.5	Nutzerverwaltung und Systemübersicht .....	124
5.5	Zusammenfassung und Erläuterung des Fortgangs der Arbeit .....	127

---

<b>6</b>	<b>Prototypische Realisierung eines WSBS</b> .....	<b>128</b>
6.1	Systematisierung der prototypischen Realisierung.....	128
6.2	Funktionsweise des WSBS .....	128
6.2.1	Systematisierung der Funktionsweise des WSBS .....	128
6.2.2	WSBS-Crawler .....	129
6.2.3	WSBS-Curator.....	133
6.3	Bedienoberflächen .....	136
6.3.1	Systematisierung der Bedienoberflächen.....	136
6.3.2	Bedienoberflächen zum Erfassen einer WSB .....	138
6.3.3	Bedienoberflächen zum Bewahren einer WSB.....	145
6.3.4	Bedienoberflächen zum Bereitstellen einer WSB .....	150
6.3.5	Weiteres Anwendungsbeispiel .....	156
<b>7</b>	<b>Zusammenfassung, Bewertung und Ausblick</b> .....	<b>159</b>
	Literaturverzeichnis .....	VII

## Abbildungsverzeichnis

	Seite
Abb. 1: Sicherungszeitpunkte der Startseite des FB02 der JLU-Gießen .....	7
Abb. 2: Schematische Darstellung des Aufbaus der Arbeit .....	14
Abb. 3: Strukturmodell: Aufgaben und Module des WSBS .....	27
Abb. 4: Erfassungsmodule im Strukturmodell des WSBS.....	32
Abb. 5: Bewahren im Strukturmodell des WSBS.....	34
Abb. 6: Bereitstellen im Strukturmodell des WSBS.....	36
Abb. 7: Signifikante Eigenschaften einer Web Site.....	38
Abb. 8: Web Site der Suchmaschine von Google.....	40
Abb. 9: Online-Konfigurator der Audi AG .....	41
Abb. 10: Anmeldeseite auf amazon.de.....	42
Abb. 11: Menü zum Verwalten von personalisierten Inhalten auf amazon.de.....	45
Abb. 12: Gestaltungsansätze der BMW-Web-Site nach Endgeräten.....	47
Abb. 13: Visualisierung der Strukturen einer Web Site.....	48
Abb. 14: Systematisierung der fachlichen Anforderungen .....	57
Abb. 15: Qualitätsanforderungen an ein klassisches AWS.....	74
Abb. 16: Vorgehensweisen des Systementwurfs.....	81
Abb. 17: Systembausteine eines WSBS .....	85
Abb. 18: Systembausteine der Datenhaltungsschicht .....	87
Abb. 19: Anwendungen zur Steuerung der Systemlösung.....	89
Abb. 20: Einordnung der Crawler-Anwendung im Strukturmodell .....	90
Abb. 21: Einordnung der Curator-Anwendung im Strukturmodell.....	91
Abb. 22: Systembaustein der Präsentationsschicht im WSBS .....	93
Abb. 23: Nutzungsoberfläche zum Erleben der WSB: Klassisch.....	94
Abb. 24: Nutzungsoberfläche zum Erleben der WSB: Slide-Show .....	95
Abb. 25: Nutzungsoberfläche zum Erleben der WSB: Dreidimensional.....	96
Abb. 26: Administrationsoberfläche .....	97
Abb. 27: Verfügbarkeitsdienste .....	98

---

Abb. 28: Abbilder und Metadaten im Strukturmodell .....	100
Abb. 29: Schematischer Aufbau eines Abbildes im WSBS .....	102
Abb. 30: Darstellung der Dateiverknüpfungen.....	103
Abb. 31: Schematische Darstellung isolierter Abbilder.....	104
Abb. 32: Mehrfachzuordnung von Dateien im WSBS.....	105
Abb. 33: Module des WSBS.....	111
Abb. 34: Module zum Erfassen im WSBS.....	112
Abb. 35: Module und Funktionen zur Erfassung im WSBS (1/2).....	113
Abb. 36: Module und Funktion zur Erfassung im WSBS (2/2) .....	115
Abb. 37: Module zum Bewahren im WSBS .....	117
Abb. 38: Module und Funktionen zum Erschließen .....	118
Abb. 39: Module und Funktionen zum Erhalten .....	120
Abb. 40: Modul zum Bereitstellen im WSBS .....	121
Abb. 41: Module und Funktionen zum Bereitstellen der WSB.....	122
Abb. 42: Positionierung der Metadaten eines Abbildes im Web Browser.....	123
Abb. 43: Module zur Kontrolle des WSBS.....	124
Abb. 44: Module und Funktionen zur Systemverwaltung (1/2).....	125
Abb. 45: Module und Funktionen zur Systemverwaltung (2/2).....	126
Abb. 46. Abschließende Darstellung des Strukturmodells.....	127
Abb. 47: Anwendungskomponenten zur Steuerung des WSBS.....	129
Abb. 48: Erfassungsprozess und Komponenten des WSBS-Crawlers .....	131
Abb. 49: Bewahren der Abbilder mit dem WSBS-Curator.....	133
Abb. 50: Bereitstellen der Abbilder mit dem WSBS-Curator .....	135
Abb. 51: Module des WSBS.....	137
Abb. 52: Eingabemaske zum Erfassen eines Adressraums.....	139
Abb. 53: Adressräume im WSBS .....	140
Abb. 54: Manuelle Sammlung von Web-Seiten im WSBS.....	141
Abb. 55: Fortschrittsanzeige manueller Sammlungen .....	142
Abb. 56: Abbild der Startseite des FB02 der JLU-Gießen am 21.06.2020.....	142

---

Abb. 57: Eingabemaske „Automatische Sammlung starten“ .....	143
Abb. 58: Sammlung starten und stoppen.....	144
Abb. 59: Technische IT-Infrastruktur im WSBS verwalten.....	146
Abb. 60: Dokumentation von Mitarbeitern im WSBS.....	147
Abb. 61: Abbilder suchen und verwalten.....	148
Abb. 62: Verwalten von Abbildern im WSBS .....	149
Abb. 63: Konfiguration der Web-Oberfläche zum Erleben der WSB .....	150
Abb. 64: Abbild einer Web-Seite des FB02 mit 1200 x 750 Pixel (1/2) .....	151
Abb. 65: Kopfzeile Reproduktionsumgebung.....	151
Abb. 66: Abbild einer Web-Seite des FB02 mit 1200 x 750 Pixel (2/2) .....	152
Abb. 67: Ausschnitt der Kalenderansicht.....	153
Abb. 68: Kopfzeile Kalenderansicht.....	153
Abb. 69: Internet-Archiv: Startseite des FB02 der JLU-Gießen im Jahr 2020 .....	154
Abb. 70: Demo-Abbild von <a href="https://www.bmw.de">https://www.bmw.de</a> , 01.08.2020 .....	156
Abb. 71: Demo-Abbild von <a href="https://www.bmw.de/neufahrzeuge">https://www.bmw.de/neufahrzeuge</a> , 01.08.2020.....	157
Abb. 72: Demo-Abbild einer Web-Seite mit 375 x 812 Pixel.....	158

## Abkürzungsverzeichnis

AWS	Anwendungssystem
BOA	Baden-Württembergisches Online-Archiv
BSB	Bayerische Staatsbibliothek
CD	Compact Disc
CRM	Customer Relationship Management
DCC	Digital Curation Centre
DIN	Deutsches Institut für Normung
DNB	Deutsche Nationalbibliothek
DSA	Data Seal of Approval
EFAC	European Framework for Audit and Certification
EU-DSGVO	Europäische Datenschutz-Grundverordnung
ERP	Enterprise Resource Planning
GB	Gigabyte
GIF	Graphics Interchange Format
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HWB	Headless Web Browser
IP	Internet Protocol
ISO	International Organization for Standardization
JPEG	Joint Photographic Experts Group
JS	JavaScript
JSON	JavaScript Object Notation
MB	Megabyte
OAIS	Open Archival Information System
PNG	Portable Network Graphics
RAR	Roshal Archive
TCP	Transmission Control Protocol
UNESCO	United Nations Educational, Scientific and Cultural Organization
URL	Uniform Ressource Locator
WARC	WebARChive
WCM	Web Content Management
WCMS	Web-Content-Management-System
WSB	Web-Site-Biographie
WSBS	Web-Site-Biographie-System
WWW	World Wide Web
XML	Extensible Markup Language

# 1 Problemstellung, Ziel und Aufbau der Arbeit

## 1.1 Problemstellung und Motivation der Arbeit

Es ist eine übliche Praxis, dass Unternehmen, Vereine oder Behörden in Schilderungen über die eigene Vergangenheit investieren.<sup>1</sup> Teilweise werden beträchtliche Summen für die professionelle Aufarbeitung und Präsentation von Unternehmens- bzw. Organisationshistorien bereitgestellt. Neben Ausstellungen oder Museen hat die Entwicklung historischer Narrationen in Buchform einen festen Platz in den Maßnahmenkatalogen von Unternehmen wie Microsoft, BMW oder Siemens.<sup>2</sup> Die Anlässe für diese Form der Geschichtsschreibung sind vielfältig. Beispielhaft lassen sich Jubiläumsfeiern, das Erbringen relevanter, gesellschaftspolitischer Leistungen, Produkteinführungen, Pionierleistungen, Erfindungen oder Börsengänge nennen.<sup>3</sup>

Charakteristisch für die derzeit übliche Ausgestaltung von Organisationshistorien ist die Exklusion des digitalen Raums. Obwohl das Internet und seine Dienste – insbesondere das World Wide Web (WWW oder Web) – zweifelsfrei markante Bestandteile der jüngeren Vergangenheit sind, stellt die historische Auseinandersetzung mit der organisationalen Web Präsenz ein Phänomen dar.<sup>4</sup> Vielmehr stehen Bemühungen im Vordergrund, die eigene Web Site dauerhaft attraktiv zu gestalten, um Aufmerksamkeit zu erregen und den andauernden technologischen Entwicklungen zu folgen.<sup>5</sup> Historische Ausarbeitungen, die über die Entwicklung der Organisation im Web berichten, sind dagegen kaum feststellbar. Dabei fungiert die Web Site eines Unternehmens nicht selten als zentrales Kommunikationsinstrument im digitalen Wirtschaftsgefüge und bildet Informationsmengen ab, deren historische Auswertung durchaus Relevanz hat.

Web Sites werden längst nicht mehr nur eingesetzt, um statische Inhalte über Produkte oder Dienstleistungen weltweit und Tageszeit-unabhängig bereitzustellen. Die andau-

---

1 Vgl. Berghoff, Hartmut: *Moderne Unternehmensgeschichte: eine themen- und theorieorientierte Einführung*, Paderborn: Schöningh 2004, S. 359.

2 Vgl. Große-Oetringhaus, Wigand F.: *Strategische Identität — Orientierung im Wandel*, Berlin, Heidelberg: Springer Berlin Heidelberg 1996, S. 6 oder Schug, Alexander: *History marketing: ein Leitfaden zum Umgang mit Geschichte in Unternehmen*, 2., unveränd. Aufl., Bielefeld: transcript 2013, S. 81.

3 Vgl. Schug, Alexander: *History marketing: ein Leitfaden zum Umgang mit Geschichte in Unternehmen*, a. a. O., S. 27 ff.

4 Vgl. Brügger, Niels: *Web historiography and Internet Studies: Challenges and perspectives*, in: *new media & society*, 15/2012, S. 3 f. oder Brügger, Niels: *Web History, an Emerging Field of Study*, in: *Web History*, Hrsg.: Brügger, Niels, New York: Peter Lang Publishing 2010, S. 8 ff.

5 Vgl. Schüle, Hubert: *Die Usability von Websites*, in: *wisu - Das Wirtschaftsstudium*, 36/2007, S. 1092.

ernde und zunehmende Nutzung des Internets geht mit einer Ausweitung digitaler Angebote einher.<sup>6</sup> Unternehmen bieten umfangreiche Funktionalitäten im Web an. Die Implementierung von Shop-Systemen oder die Bereitstellung interaktiver, kollaborativer, Workflow-basierter oder personalisierter Dienste ist längst allgegenwärtig.<sup>7</sup> Zudem werden geschäftsrelevante IT-Systeme unmittelbar mit der Web Site verbunden. Geschäftsmodelle, die im Wesentlichen auf digitalen Angeboten fußen, sind keine Seltenheit mehr. Heute prägen Unternehmen wie „Google“ oder „Amazon“ das alltägliche Leben in unterschiedlichsten Bereichen. Die Web Site fungiert für viele Organisationen längst als ein integriertes und zentrales Instrument zur Abwicklung wertschöpfender Aktivitäten.

Betrachtet man die historische Entwicklung des von Tim Berners-Lee im Jahre 1989 entwickelten World Wide Web, kann diesem System zweifelsfrei ein nachhaltiger Einfluss auf die Gesellschaft und die darin verwurzelten Unternehmen attestiert werden.<sup>8</sup> Eine von der UNESCO bereits im Jahre 2003 verabschiedete „Charta zur Bewahrung des digitalen Kulturerbes“ – und damit auch zur Bewahrung von Web Sites – unterstreicht diesen Einfluss längst.<sup>9</sup> Zum Status Quo lassen sich jedoch nur wenige Forschungsarbeiten erkennen, die sich aus historisch motivierten Fragestellungen mit dem technischen Konstrukt „Web“ oder einzelnen Bestandteilen dieses Konstrukts – im Speziellen Web Sites – auseinandersetzen. Insbesondere im Kontext von Unternehmen bzw. Organisationen sind Ausarbeitungen, die sich mit historischen Zuständen und Veränderungen einer Web Site befassen, kaum oder nur schwer feststellbar.<sup>10</sup> Probleme beim Erstellen entsprechender Schilderungen resultieren unter anderem aus dem Fehlen konsensfähiger theoretischer Konzepte und dem Mangel geeigneter Bezugssysteme zur Erfassung, Bewahrung und Nutzbarmachung relevanter Quellen aus dem World Wide Web.<sup>11</sup>

---

6 Vgl. eMarketer (Hrsg.): eMarketer Updates Worldwide Internet and Mobile User Figures, Online im Internet: <https://www.emarketer.com/Article/eMarketer-Updates-Worldwide-Internet-Mobile-User-Figures/1015770>, 25.05.2017.

7 Vgl. Kappel, Gerti; Pröll, Birgit; Reich, Siegfried; Retschitzegger, Werner: Web Engineering – Die Disziplin zur systematischen Entwicklung von Web-Anwendungen, in: Web engineering: systematische Entwicklung von Web-Anwendungen, Hrsg.: Kappel, Gerti, 1. Aufl, Heidelberg: dpunkt-Verl 2004, S. 7.

8 Vgl. Berners-Lee, Tim; Döpfner, Matthias: Vage, aber HOCHINTERESSANT, in: WELT AM SONNTAG, 2017, S. 11 ff.

9 Vgl. UNESCO (Hrsg.): Charta zur Bewahrung des digitalen Kulturerbes, Online im Internet: <http://www.unesco.de/infothek/dokumente/unesco-erklarungen/charta-zur-bewahrung-des-digitalen-kulturerbes.html>, 26.05.2017.

10 Eine Übersicht relevanter Arbeiten stellt Brügger dar. Vgl. Brügger, Niels: Web History, an Emerging Field of Study, a. a. O., S. 8 ff.

11 Vgl. Brügger, Niels: When the present web is later the past: Web historiography, digital history, and Internet Studies, in: Historical Social Research/Historische Sozialforschung, 2012, S. 105 oder Brügger, Niels: Web History, an Emerging Field of Study, a. a. O., S. 6 ff. oder Brügger, Niels: Web historiography and Internet Studies: Challenges and perspectives, a. a. O., S. 4 ff.

Die Rolle als zentrales Bezugssystem zur Erschließung historisch relevanter Quellen hat seit jeher das Archiv inne.<sup>12</sup> Sowohl für Behörden, die eine Archivierung aufbewahrungswürdiger Dokumente mit Hilfe der Staatsarchive umsetzen, als auch für andere Institutionen existieren diverse Vorgaben, die aus juristischen, politischen oder auch kulturellen Gründen Archivierungsaufgaben bedingen. Archive, die mit der Aufbewahrung geschäftsrelevanter Materialien betraut sind, haben längst einen festen Platz in organisationalen Prozessen und Strukturen. Gesetzliche Forderungen, das Risiko der Produkthaftung oder ein internes Qualitätsmanagement sind einige der Treiber, die eine Implementierung entsprechender Archivsysteme bedingen.<sup>13</sup> Eine Ausnahme stellt bis heute die historische Archivierung einer Web Site dar. Sind seit einigen Jahren auf nationaler und internationaler Ebene Initiativen zu erkennen, die sich um die nachhaltige Bewahrung des Kulturgutes „Web“ bemühen, ist bei Unternehmen, Vereinen oder Behörden ein vergleichbares Engagement selten zu beobachten.<sup>14</sup>

Auch wenn die Konzepte des Web Content Managements eine Archivierung im Lebenszyklus der Web Contents vorsehen, sind Web-Content-Management-Systeme (WCMS), die häufig im Organisationskontext zur Bereitstellung einer Web Site eingesetzt werden, für eine langfristige Speicherung und Reproduktion früherer Zustände einer Web Site kaum geeignet.<sup>15</sup> Bereits der Umstand, dass im Laufe der Zeit unterschiedliche Hersteller von WCMS mit differierenden Produktlösungen und Schnittstellen herangezogen werden können, bedingt mögliche Migrationsprozesse, die mit einem Content-Verlust verbunden sind. Restdaten, die nicht migriert werden können, müssen dann z. B. auf geeignete Datenträger transferiert und für die Nachwelt dokumentiert werden, um nicht gänzlich in Vergessenheit zu geraten.<sup>16</sup> Weiterhin ist Web Content durch eine eigentümliche Flüchtigkeit charakterisiert. Was heute im Internet veröffentlicht wurde, kann morgen verschwunden sein. Im Gegensatz zu Büchern, deren Überarbeitung mit einer neuen, gedruckten oder digitalen Ausgabe verbunden ist, sind Inhalte im Web ersetzbar, ohne dass

---

12 Vgl. Nentwich, Michael: *Cyberscience: research in the age of the internet*, Vienna: Austrian Acad. of Sciences Press 2003, S. 38.

13 Vgl. Schneider, Holger: *Digitale Amnesie: Langzeitarchivierung digitaler Dokumente im betrieblichen Umfeld*, Norderstedt: Books on Demand 2012, S. 22 ff.

14 Vgl. Haber, Peter: *Digital past: Geschichtswissenschaft im digitalen Zeitalter*, München: Oldenbourg Verlag 2011, S. 60 oder Masanès, Julien: *Web archiving*, Berlin, London: Springer 2010, S. 41.

15 Vgl. Krüger, Jörg Dennis; Kopp, Matthias: *Web-Content managen: professioneller Einsatz von Content-Management-Systemen*, München: Markt-und-Technik-Verlag 2002, S. 54 ff.

16 Vgl. Jochemich, Hanno: „Fehler HTTP 404 – Seite nicht gefunden“ Oder: Zum Stand der Archivierung von Webinhalten in den öffentlich-rechtlichen Rundfunkanstalten, in: *Das Gedächtnis des Rundfunks: die Archive der öffentlich-rechtlichen Sender und ihre Bedeutung für die Forschung*, Hrsg.: Hasselbring, Bettina; Bernard, Birgit; Behmer, Markus, Wiesbaden: Springer VS 2014, S. 149 f.

eine zusätzliche Version entsteht.<sup>17</sup> Um den Datenverlust durch Flüchtigkeit oder die begrenzte Einsatzdauer eines WCMS zu kompensieren, sind originäre, zusätzliche IT-Systeme nötig, die die vollständige Bewahrung der Vergangenheit des Web-Angebots einer Organisation sicherstellen. Systemlösungen, die umfassend auf die beschriebene Problematik reagieren, sind innerhalb einzelner Organisationen schwer zu finden. Es fehlen Funktionalitäten, um die Inhalte einer Web Site aus einer fachlichen bzw. inhaltlichen, technischen und organisatorischen Sicht in geeigneter Weise abzubilden, zu speichern und für historische Ausarbeitungen bereitzustellen.<sup>18</sup>

Diese Problematik lässt sich nahezu ungefiltert auf die Entstehung von Web-Archiven nationaler Bibliotheken und international agierender Institutionen transportieren.<sup>19</sup> Als Vorreiter international agierender Institutionen kann das „Internet-Archiv“ angeführt werden. Bereits im Jahre 1996 wird es als erstes digitales Gedächtnis zur Sammlung und Aufbewahrung von Publikationen aus dem World Wide Web ins Leben gerufen. Brewster Kahle, der Gründer des Internet-Archivs, verfolgte den Gedanken, in einem zentralen Archiv veröffentlichte Web Sites für die Nachwelt zu sichern. Das drohende Verschwinden von Texten und Bildern im Web sollte damit eingedämmt und der Verlust digitaler Erbgüter verhindert werden. In Deutschland hat sich ein vergleichbares Engagement erst in den letzten Jahren gezeigt. Projekte, die im Bereich der nationalen Bibliotheken eine Bewahrung des Kulturgutes Web vorantreiben, sind z. B. das Web-Archiv „edoweb“, das Baden-Württembergische Online-Archiv (BOA) oder die Web-Archivierung der Bayerischen Staatsbibliothek (BSB). Zudem wurde der Sammelauftrag der Deutschen Nationalbibliothek (DNB) durch den Gesetzgeber auf im Web veröffentlichte Publikationen ausgeweitet.<sup>20</sup> Von diesem Auftrag sind seit dem Jahr 2013 ca. 700 Web Sites betroffen.<sup>21</sup>

Trotz der Spiegelung des World Wide Web durch nationale Bibliotheken oder international agierende Institutionen besteht Handlungsbedarf, besonders mit Blick auf die historische Auswertung und vollständige Reproduktion von Unternehmens-Web-Sites.<sup>22</sup>

---

17 Vgl. Masanès, Julien: Web archiving, a. a. O., S. 12 ff.

18 Vgl. Jochemich, Hanno: „Fehler HTTP 404 – Seite nicht gefunden“ Oder: Zum Stand der Archivierung von Webinhalten in den öffentlich-rechtlichen Rundfunkanstalten, a. a. O., S. 420 ff.

19 Ein weltweiter Überblick über Archivierungsinitiativen findet sich in Schwenck, Rebecca: Langzeitarchivierung digitaler Informationen: Ein internationaler Überblick über Organisation und Projektarbeit, Saarbrücken: VDM 2008, S. 26 ff.

20 Vgl. Bundesministerium der Justiz und für Verbraucherschutz (Hrsg.): Gesetz über die Deutsche Nationalbibliothek (DNBG) § 3 Medienwerke, Online im Internet: [https://www.gesetze-im-internet.de/dnbg/\\_3.html](https://www.gesetze-im-internet.de/dnbg/_3.html), 26.05.2017.

21 Vgl. Beiner, Tobias; Hagenah, Ulrich; Kugler, Anna: Es war einmal eine Website... - Kooperative Webarchivierung in der Praxis, in: o-bib, 1/2014, S. 2 f.

22 Vgl. Hale, Scott A.; Blank, Grant; Alexander, Victoria D.: Live versus archive: Comparing a web archive to a population of web pages, in: The web as history: using web archives to understand the past and the present, Hrsg.: Brügger, Niels; Schroeder, Ralph, London: UCL Press 2017, S. 45 ff.

Abbilder einzelner Web Sites sind in den Sammlungen der Web-Archive nicht selten fehlerhaft, unvollständig oder fehlen völlig.<sup>23</sup> Beim Übertragen einer Web Site in die Infrastruktur eines Web-Archivs ist ein Verlust von Bildern, Texten, Funktionen oder anderen Inhalten symptomatisch.<sup>24</sup> Zudem erschwert ein systematischer Mangel an geeigneten Kontext-Informationen, die zur Interpretation und Nachvollziehbarkeit früherer Zustände einer Web Site benötigt werden, die historische Arbeit mit verfügbaren Beständen. Betrachtet man die Geschichte einer Web Site als Produkt, das Kausalitäten und Zusammenhänge beschreibt, sind diese Informationsmengen grundlegend, um die Spuren der Vergangenheit zu deuten und eine fundierte, historische Ausarbeitung zu formen.<sup>25</sup>

Neben dem Mangel von Kontext-Informationen und Defiziten bei der Überführung einer Web Site in eine geeignete Speicherumgebung stellt Flüchtigkeit, kombiniert mit der kontinuierlich wachsenden Menge an Inhalten, eine wesentliche Herausforderung dar. Das Internet-Archiv umfasst bis dato eine der umfangreichsten öffentlichen Sammlungen von Angeboten aus dem World Wide Web (WWW).<sup>26</sup> Bereits im Jahre 2015 wird das digitale Gedächtnis des Internet-Archivs mit mehr als einer Milliarde Web-Seiten pro Woche gespeist und es ist anzunehmen, dass diese Sammelfrequenz in den letzten Jahren nicht abgenommen hat. Trotz der wöchentlich gesicherten Menge an Kopien einzelner Seiten des Webs, liegt ein „Sieg“ im Kampf gegen das Vergessen im digitalen Raum – die sogenannte „digitale Amnesie“ – noch in unbestimmter Ferne.

Eine Vielzahl an Inhalten, die im WWW veröffentlicht werden, verändern sich in Intervallen, die zum gegenwärtigen Zeitpunkt weder aus organisatorischer noch aus technischer Sicht durch eine zentrale Organisation – wie das Internet-Archiv – erfasst werden können. Beispielhaft lassen sich an dieser Stelle unter anderem stark dynamische, auf Nutzereingaben basierende Inhalte anführen.<sup>27</sup> Aber auch Web Sites, die eine geringe Aktualisierungshäufigkeit aufweisen, sind auf Grund der schieren Inhaltsmenge im Web nicht selten lückenhaft in den digitalen Sammlungen der Web-Archive vertreten.

---

23 Vgl. z. B. Masanès, Julien: Web archiving, a. a. O., S. 20 oder Brown, Adrian: Archiving websites: a practical guide for information management professionals, London: Facet Pub 2006, S. 28 oder Brügger, Niels: Web historiography and Internet Studies: Challenges and perspectives, a. a. O., S. 5 f. oder Haber, Peter: Digital past: Geschichtswissenschaft im digitalen Zeitalter, a. a. O., S. 61.

24 Vgl. Masanès, Julien: Web archiving, a. a. O., S. 15 oder Hessischer Rundfunk (Hrsg.): Gegen den Gedächtnisverlust: Das Archiv im digitalen Zeitalter, Online im Internet: <http://www.hr-inforadio.de/podcast/wissen/podcast-episode-5084.html>, 05.05.2017.

25 Vgl. Brügger, Niels: Web History, an Emerging Field of Study, a. a. O., S. 5.

26 Vgl. Ainsworth, Scott; AlSum, Ahmed; SalahEldeen, Hany; Weigle, Michele C.; Nelson, Michael L.: How much of the Web is Archived?, Online im Internet: <http://arxiv.org/abs/1212.6177>, 15.05.2016.

27 Vgl. Lobe, Adrian: Wir sammeln eine Milliarde Seiten pro Woche, in: Frankfurter Allgemeine Zeitung (F.A.Z.), 121/2015, S. 13.

Mehr als 50 Prozent der Weltbevölkerung nutzen das Internet und seine Dienste.<sup>28</sup> Kontinuierlich werden neue Inhalte publiziert und bereits veröffentlichte Inhalte verändert oder gelöscht. Das Bestreben, sämtliche Zustände und Veränderungen dieser Inhalte in einem System abzubilden, erscheint zum gegenwärtigen Zeitpunkt illusorisch. Archivierungsstrategien umfassen notwendigerweise Priorisierungen, die aus politischer, kultureller oder anderweitiger Brisanz definiert werden, um vermeintlich relevante Informationen aus den Datenmeeren des Web zu kopieren.<sup>29</sup> Diese Vorgehensweise bedingt, dass Web Sites, deren Informationsmengen nicht die nötige Brisanz aufweisen, mit einer geringeren Häufigkeit oder überhaupt nicht erfasst und langfristig gesichert werden.<sup>30</sup>

Ergänzend dazu ist es möglich, Schlüsselwörter in den Quellcode einer Web Site einzubinden und damit die Erfassung durch Sammelwerkzeuge der Web-Archive zu vermeiden. Diese Möglichkeit der aktiven Exklusion einer Web Site ist Bestandteil einer geführten Debatte. Auslöser ist ein Vorhaben des Internet-Archivs in San Francisco: Es sieht vor, diese Schlüsselwörter im Rahmen der kontinuierlichen Archivierungsprozesse zu ignorieren. Alle erfassten Seiten sollen, unabhängig von entsprechend definierten Schlüsselwörtern, abgebildet und im Archiv gesichert werden. Der Bestand des Internet-Archivs kann so ausgeweitet und die Bewahrung des digitalen Kulturgutes „Web“ vorangetrieben werden.<sup>31</sup> Rechtliche Fragestellungen sind in diesem Zusammenhang kritisch zu betrachten, da die Sammelwerkzeuge Angaben im Web-Site-Quellcode ignorieren und Unternehmenseigentum kopieren.

Weiterhin ist auf Ebene der Web-Site-Betreiber keine flächendeckende Akzeptanz für kulturell bzw. historisch motivierte Sicherungsmaßnahmen der eigenen Web Site zu erkennen. Gründe lassen sich im Interesse an der ausschließlichen Verfügungsgewalt über die eigenen Inhalte oder der Vermeidung potentieller Vervielfältigungen urheberrechtlich geschützter Elemente vermuten.<sup>32</sup> Mit Bezug zu geeigneten rechtlichen Rahmenbedingungen ist z. B. in Deutschland noch kein konsensfähiges Ergebnis feststellbar.<sup>33</sup> Es ist

---

28 Vgl. Berners-Lee, Tim; Döpfner, Matthias: Vage, aber HOCHINTERESSANT, a. a. O., S. 11 ff. und Statista (Hrsg.): Number of internet users worldwide from 2005 to 2018 (in millions), 29.04.2019 Online im Internet: <https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/>, 29.04.2019.

29 Vgl. Brown, Adrian: Archiving websites: a practical guide for information management professionals, a. a. O., S. 26 ff.

30 Vgl. Lobe, Adrian: Wir sammeln eine Milliarde Seiten pro Woche, a. a. O., S. 13.

31 Vgl. Graham, Mark: Robots.txt meant for search engines don't work well for web archives, Online im Internet: <https://blog.archive.org/2017/04/17/robots-txt-meant-for-search-engines-dont-work-well-for-web-archives/>, 05.05.2017.

32 Vgl. Deutscher Bundestag (Hrsg.): Entwurf eines Gesetzes über die Deutsche Nationalbibliothek (DNBG), Online im Internet: <http://dipbt.bundestag.de/dip21/btd/16/003/1600322.pdf>, 05.05.2017.

33 Vgl. Beiner, Tobias; Hagenah, Ulrich; Kugler, Anna: Es war einmal eine Website... - Kooperative Webarchivierung in der Praxis, a. a. O., S. 294.

daher anzunehmen, dass die Evolution einer Vielzahl von Web Sites unzuverlässig und lückenhaft in digitalen Gedächtnissen der Web-Archive, z. B. in den Sammlungen des Internet-Archivs, abgebildet wird.<sup>34</sup>

Exemplarisch zeigt Abbildung 1 eine Übersicht gesicherter Kopien der Startseite des Fachbereichs 02, der Justus-Liebig-Universität Gießen.<sup>35</sup> Im Jahr 2016 wurden an 21 Tagen, dargestellt durch kreisförmige Markierungen in der kalendarischen Übersicht, Kopien der Startseite (<http://www.uni-giessen.de/home/fb02/>) gesichert und bereitgestellt.

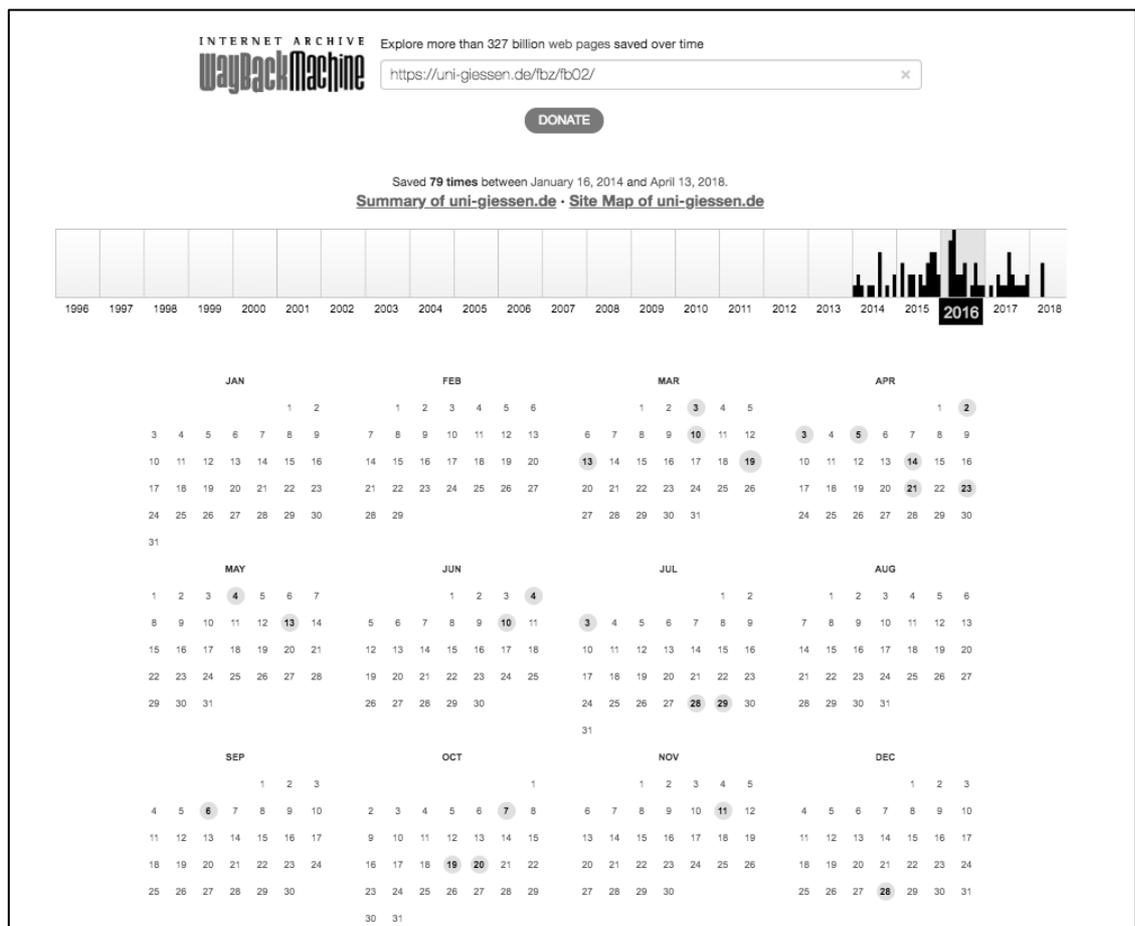


Abb. 1: Sicherungszeitpunkte der Startseite des FB02 der JLU-Gießen

Auf Basis der in Abbildung 1 dargestellten Sicherungszeitpunkte lässt sich eine überschaubare Menge an Zuständen rekonstruieren. Eine vollständige Reproduktion sämtlicher Veränderungen (z. B. täglich veröffentlichte News-Board-Einträge) dieser spezifischen Web-Seite ist mit Hilfe des Internet-Archivs nicht möglich. Weiter sind die gesicherten Kopien das Ergebnis einer randomisierten bzw. randomisiert wirkenden Auswahl

34 Vgl. Lobe, Adrian: Wir sammeln eine Milliarde Seiten pro Woche, a. a. O., S. 13.

35 Vgl. Internet Archive (Hrsg.): Internet Archive Wayback Machine, Online im Internet: [http://web.archive.org/web/20161119154605\\*/http://wiwi.uni-giessen.de/home/fb02/](http://web.archive.org/web/20161119154605*/http://wiwi.uni-giessen.de/home/fb02/), 28.03.2017.

an Sicherungszeitpunkten. Eine Garantie dafür, dass alle benötigten Inhalte für eine historische Schilderung erfasst wurden, ist nicht gegeben.

Der Verlust digitaler Daten und die damit verbundene digitale Amnesie sind jedoch nicht nur mit Bezug zur Erfassung und Erschließung relevanter Inhalte herausfordernd, sondern auch bei der Auswahl und Verwendung geeigneter Speichertechniken. Das Internet und seine Dienste stehen im Mittelpunkt täglicher Anwendungen. Kulturgüter werden in Form von digitalen Bildern, Texten, Fotos, Videos, E-Mails, Spielen oder Programmen erzeugt und über das Netzwerk „Internet“ zum Abruf bereitgestellt.<sup>36</sup> Gleichlaufend ist die Transformation analoger Artefakte in den digitalen Raum allgegenwärtig. Längst haben unter anderem Museen ihre analogen Ausstellungsstücke in die virtuelle Welt transferiert, um diese von beliebigen Orten, auf unterschiedlichsten Endgeräten zugänglich zu machen und damit dem Trend der Digitalisierung Rechnung zu tragen.<sup>37</sup> Eine Lösung, um die Vielzahl digitaler Erb- und Kulturstücke langfristig für die Nachwelt zu sichern, steht jedoch noch aus. Neben rechtlichen Rahmenbedingungen und Problemen bei der Erfassung, stellt sich vor allem die Frage nach geeigneten, technischen Lösungen zur langfristigen Sicherung der Haltbarkeit und Lesbarkeit von digitalen Daten auf einem bzw. wechselnden Datenträgern.<sup>38</sup>

Im Vergleich mit über Jahrhunderte bewährten Medien wie z. B. Papier haben digitale Datenträger wie CDs eine geringe Haltbarkeit und unterliegen einem schnellen technologischen Wandel.<sup>39</sup> Die Substanzerhaltung binärer Daten ist daher an einen kontinuierlichen Prozess aus emulieren oder migrieren gebunden. Anerkannte Konzepte und geeignete Testverfahren werden in diesem Zusammenhang noch vermisst. Die Gefahr des Verlusts kulturell relevanter digitaler Artefakte ist weiterhin akut.<sup>40</sup>

Bereits an dieser Stelle kann konstatiert werden, dass retrospektive Ansätze, die vollständige Evolutions- bzw. Lebensgeschichte einer Web Site zu dokumentieren, unmittelbar durch die oben genannten Probleme erschwert bzw. verhindert werden.

---

36 Vgl. Gesellschaft für Informatik e. V. (Hrsg.): Digitales Kulturerbe, Online im Internet: <https://www.gi.de/themen/grand-challenges/digitale-kultur.html>, 28.03.2017.

37 Vgl. Zschunke, Peter; Schmidt, Julia: Digitalisierung: Museen sind bereit zu teilen, Online im Internet: <https://www.heise.de/newsticker/meldung/Digitalisierung-Museen-sind-bereit-zu-teilen-3609717.html>, 28.03.2017.

38 Vgl. Beiner, Tobias; Hagenah, Ulrich; Kugler, Anna: Es war einmal eine Website... - Kooperative Webarchivierung in der Praxis, a. a. O., S. 291.

39 Vgl. Schwenck, Rebecca: Langzeitarchivierung digitaler Informationen: Ein internationaler Überblick über Organisation und Projektarbeit, a. a. O., S. 15 f. oder Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, Saarbrücken: AV Akademikerverlag 2012, S. 23 ff.

40 Vgl. Gesellschaft für Informatik e. V. (Hrsg.): Digitales Kulturerbe, a. a. O., 28.03.2017.

Gedächtnisinstitute wie das Internet-Archiv oder die Deutsche Nationalbibliothek stellen in ihren Sammlungen lückenhaft Zustände einzelner Web Sites aus dem World Wide Web sicher. Unversehrte Abbilder, die sämtliche Veränderungen einer spezifischen Web Site aufweisen, sind in den gegenwärtigen Systemlösungen nicht reproduzierbar. Ähnlich kritisch stellt sich die Situation mit Sicht auf die organisationsinternen Sicherungssysteme dar.

Web-Content-Management-Systeme, die häufig zum Betrieb einer Unternehmens-Web-Site verwendet werden, bieten unter Umständen Archivierungsfunktionen an. Die Reproduktion veralteter bzw. inaktiver Seiten ist systemseitig jedoch häufig nicht mehr realisierbar, wenn entsprechende Systeme über die Zeit ersetzt, erweitert oder kombiniert werden. Die daraus resultierende Veränderung in der IT-Anwendungslandschaft setzt aufwendige Migrationsprozesse voraus, um den Datenbestand der Web Site zu erhalten. Aussparungen führen häufig und unmittelbar zu Content-Verlust.<sup>41</sup> Als Konsequenz sind historisch motivierte Initiativen, die Lebensgeschichte einer Web Site zu dokumentieren, unbedingt mit proaktivem Handeln verbunden. Notwendige Informationen müssen aus organisationsinternen und organisationsexternen Quellen zusammengetragen werden, falls diese überhaupt vorhanden und zugänglich sind. Neben öffentlichen Sammlungen der Web-Archive sind als Quellen beispielsweise Bücher, Berichte und Protokolle in Papierform, Inhalte aus traditionellen Massenmedien oder Interviews anzuführen.<sup>42</sup>

Die Konstruktion eines IT-Systems, das auf die beschriebene Problematik reagiert, ist in der wissenschaftlichen Literatur nicht zu erkennen.<sup>43</sup> Dieser Zustand beschreibt die Problemstellung und begründet die Forschungsmotivation der vorliegenden Untersuchung. Sie widmet sich der Grundlagenbeschreibung sowie der Analyse und Definition eines IT-Systems zur nutzenstiftenden, systemseitigen Dokumentation der vollständigen Lebensgeschichte einer organisationalen Web Site, der sogenannten „*Web-Site-Biographie*“. Eine gesonderte Definition des Begriffs erfolgt in Kapitel 2.2 der vorliegenden Arbeit.

Im folgenden Kapitel wird nun zunächst der Untersuchungsbereich der vorliegenden Arbeit definiert und deren Zielsetzung konkretisiert.

---

41 Vgl. Jochemich, Hanno: „Fehler HTTP 404 – Seite nicht gefunden“ Oder: Zum Stand der Archivierung von Webinhalten in den öffentlich-rechtlichen Rundfunkanstalten, a. a. O., S. 419 f.

42 Vgl. Brügger, Niels: Web History: An Analytical Grid, in: Web History, Hrsg.: Brügger, Niels, New York: Peter Lang Publishing 2010, S. 24. oder Brügger, Niels: Web historiography and Internet Studies: Challenges and perspectives, a. a. O., S. 756.

43 Vgl. Balzert, Helmut: Lehrbuch der Software-Technik, 2. Aufl., 1. Nachdr., Heidelberg: Spektrum Akad. Verlag 2001, S. 35 f. und Gabriel, Roland; Weber, Peter; Schroer, Nadja; Lux, Thomas: Basiswissen Wirtschaftsinformatik, 2. Aufl, Dortmund: W3L-Verl 2014, S. 7.

## 1.2 Untersuchungsbereich und Zielsetzung

Der Untersuchungsbereich der vorliegenden Arbeit umfasst Organisationen aus Wirtschaft und Verwaltung, die mit einer Web Site im World Wide Web vertreten sind. In den Untersuchungsbereich eingebettet ist das Untersuchungsobjekt. Das konkrete Untersuchungsobjekt der vorliegenden Arbeit wird durch organisationale Web Sites beschrieben, die in Wirtschaft und Verwaltung zur Abwicklung täglicher Aktivitäten bzw. Geschäftstätigkeiten im digitalen Raum verwendet werden. Ein Fokus auf eine bestimmte Form von Geschäftsaktivitäten bzw. Geschäftsmodellen wird nicht gelegt.<sup>44</sup>

Das technische Konstrukt „Web Site“ wird im Rahmen der vorliegenden Untersuchung als eine eigenständig oder durch Dienstleister erstellte und betriebene Präsenz im Web verstanden, die unter einem der Organisation zugewiesenen, individuellen Adressraum präsentiert wird. Eine fachliche, organisatorische und technische Verfügungsgewalt über sämtliche Inhalts- und Gestaltungselemente wird vorausgesetzt. Durch die vorliegende Abgrenzung sollen weitere Internet-Präsenzen eines Unternehmens bzw. einer Organisation exkludiert werden, die auf Plattformen von Anbietern wie „Facebook“ oder „YouTube“ erzeugt und bereitgestellt sind. Diese ermöglichen im Regelfall keine vollständige Verfügungsgewalt, da Inhalte in plattformspezifische Vorlagen integriert und die technische Bereitstellung bzw. Ausstattung nicht beeinflusst werden kann.

Mit dem Untersuchungsobjekt verbunden ist die Zielsetzung der vorliegenden Forschungsarbeit. Die Zielsetzung wird durch das stringente Herleiten eines IT-Systems zur vollständigen Dokumentation der Lebensgeschichte einer Web Site beschrieben. Diese Zielsetzung wird durch folgende Forschungsfragen konkretisiert:

1. Welche Eigenschaften einer Web Site müssen für die vollständige und systemseitige Dokumentation der Lebensgeschichte einer Web Site berücksichtigt werden? Wie lassen sich eben diese Eigenschaften des Bezugsobjektes „Web Site“ herleiten und welche Herausforderungen sind mit diesen Eigenschaften im Kontext einer digitalen Speicherumgebung verbunden?
2. Welche Anforderungen muss das IT-System erfüllen, um die Lebensgeschichte einer Web Site vollständig – d. h. unter Berücksichtigung aller relevanter Web-Site-Eigenschaften – zu dokumentieren und in einer geeigneten technischen Umgebung nutzenstiftend wiederzugeben?

---

44 Vgl. Wirtz, Bernd W.; Becker, Daniel R.: Geschäftsmodellansätze und Geschäftsmodellvarianten im Electronic Business, in: WiSt – Wirtschaftsstudium, Heft 2/2002, S. 90.

3. Wie können die Anforderungen an ein solches Web-Site-Biographie-System technisch konkretisiert und in einem angemessenen Systementwurf systematisch skizziert werden?
4. Ist die Vollständigkeit der Anforderungen und die Anwendbarkeit der Systemlösung im Kontext der eingangs formulierten Problemstellung gegeben bzw. kann diese bestätigt werden und das IT-System in der Praxis eine Anwendung finden?

Im folgenden Kapitel wird die Forschungsmethode zur Erreichung der formulierten Zielsetzungen erläutert und der Aufbau der Arbeit skizziert.

### 1.3 Forschungsmethode und Aufbau der Arbeit

Die vorliegende Arbeit wird der Disziplin der Wirtschaftsinformatik zugeordnet. Betrachtet man die Forschungsziele der Wirtschaftsinformatik, lassen sich diese in Erkenntnis- und Gestaltungsziele gliedern. Erkenntnisziele ergeben sich aus dem Verstehen bestehender Sachverhalte. Gestaltungsziele sind durch das Schaffen neuartiger Artefakte charakterisiert.<sup>45</sup> Mit Bezug zur Zielsetzung der vorliegenden Untersuchung – dem Dokumentieren der Lebensgeschichte einer Web Site – kann diese zweifelsfrei als Gestaltungsziel klassifiziert werden. Da das Gestaltungsziel einer Forschungsarbeit nicht immer trennscharf von einem Erkenntnisziel separiert werden kann, wird neben der Schaffung eines neuartigen Artefakts ein Erkenntnisziel unterstützt. Es manifestiert sich im Verständnis für die Notwendigkeit einer langfristigen Dokumentation der Lebensgeschichte von organisationalen Web Sites und dem damit verbundenen, wirtschaftlichen und historischen Nutzen.<sup>46</sup> Das zentrale Forschungsziel bleibt jedoch gestalterischer Natur und bedingt den Einsatz einer gestaltungsorientierten Forschungsmethode. In diesem Zusammenhang wird das von Pfeffers et. al. publizierte Modell für gestaltungsorientierte Forschungsarbeiten angewandt. Der Forschungsprozess umfasst die folgenden fünf Aktivitäten: *Problemidentifikation und Motivation, Definition der Ziele, Entwurf und Entwicklung, Demonstration und Evaluation und Kommunikation.*<sup>47</sup>

---

45 Vgl. Becker, Jörg; Holten, Roland; Knackstedt, Ralf; Niehaves, Björn: Forschungsmethodische Positionierung in der Wirtschaftsinformatik – epistemologische, ontologische und linguistische Leitfragen, in: Arbeitsberichte des Instituts für Wirtschaftsinformatik, Hrsg.: Institut für Wirtschaftsinformatik, Westfälische Wilhelms-Universität Münster, Arbeitsbericht Nr. 93, Münster: 2003, S.11 ff. oder Picot, Arnold: Richtungsdiskussionen in der Wirtschaftsinformatik, in: Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung, 62/2010, S. 666 f.

46 Vgl. Becker, Jörg; Holten, Roland; Knackstedt, Ralf; Niehaves, Björn: Forschungsmethodische Positionierung in der Wirtschaftsinformatik – epistemologische, ontologische und linguistische Leitfragen, a. a. O., S.11.

47 Vgl. Pfeffers, Ken; Tuunanen, Tuure; Rothenberger, Marcus A.; Chatterjee, Samir: A design science research methodology for information systems research, in: Journal of Management Information Systems, Vol. 24 Nr. 3/2007, S. 46 ff.

1. *Problemidentifikation und Motivation:* Im ersten Schritt der Arbeit wird die zugrundeliegende Problemstellung und Forschungsmotivation hergeleitet, begründet und ausformuliert. Die Feststellung des Problems erfolgt für die vorliegende Untersuchung durch eine Analyse bestehender und verfügbarer Literatur sowie durch eine Betrachtung und Auswertung bestehender und in der Praxis angewandter IT-Systeme, die mit der Bewahrung digitaler Güter – im Speziellen Web Sites – betraut sind. Im Zusammenhang mit der Literaturanalyse werden sowohl Werke berücksichtigt, die sich mit der Relevanz von klassischen Unternehmensgeschichten befassen, als auch Veröffentlichungen, die sich mit der historischen Aufarbeitung einer organisationalen Web Site auseinandersetzen.<sup>48</sup> Zudem haben vor allem in Fachkreisen gefestigte Werke zur Web- und Web-Site-Archivierung sowie globale Standards (u. a. die DIN-Norm 31644) zur langfristigen Sicherung digitaler Güter, einen zentralen Einfluss auf die Bestimmung der Problemstellung und Forschungsmotivation.<sup>49</sup>
2. *Definition der Ziele:* Aufbauend auf der Problemstellung und Forschungsmotivation sind die Forschungsziele der Arbeit zu formulieren. Das übergeordnete Forschungsziel der Arbeit wird durch das Dokumentieren der Biographie bzw. Lebensgeschichte einer Web Site beschrieben. Es ist durch einen gestaltungsorientierten Forschungsprozess geprägt und explorativer Natur. Bei der Ausgestaltung bzw. Wahl von geeigneten Verfahren zur Zielerreichung kann grundsätzlich zwischen qualitativen und quantitativen Lösungswegen unterschieden werden.<sup>50</sup> Die Zielformulierungen in der vorliegenden Untersuchung sind der Art gewählt, dass ein qualitativer Lösungsweg zur Zielerreichung forciert wird. Dieser ist wiederum durch einen deduktiven Forschungsprozess geprägt, der sich aus den zentralen Verfahren des Software Engineerings speist. Als (Teil-)Ziele können daher auch die systematische Anforderungsanalyse, ein umfassender Systementwurf sowie eine prototypische Realisierung als Machbarkeits-Nachweis bestimmt werden.
3. *Entwurf und Entwicklung:* Der Entwurf und die Entwicklung eines IT-Systems zum Dokumentieren der Biographie einer Web Site orientieren sich – gemäß der Zielsetzung – an den klassischen Schritten des Software Engineerings. Im Speziellen kom-

---

48 Vgl. Brügger, Niels: Web History, an Emerging Field of Study, a. a. O., S. 5 ff.

49 Vgl. hierzu z. B. die Auflistung und Erläuterung relevanter Normen in Brown, Adrian: Practical digital preservation: a how-to guide for organizations of any size, London: Facet Publ 2013, S. 86.

50 Vgl. Recker, Jan: Scientific research in information systems: a beginner's guide, Heidelberg: Springer 2013, S. 35 oder Pfeffers, Ken; Tuunanen, Tuure; Rothenberger, Marcus A.; Chatterjee, Samir: A design science research methodology for information systems research, a. a. O., S. 12.

men eine *Anforderungsanalyse*, ein *Systementwurf* und die *prototypische Realisierung* zum Tragen.<sup>51</sup> Den Ausgangspunkt für den systematischen Aufbau der Anforderungsanalyse bilden signifikante Eigenschaften einer Web Site. Diese werden im Rahmen einer *Eigenschaftsanalyse* abgeleitet und bestimmen in der vorliegenden Arbeit die fachlichen, organisatorischen und technischen Anforderungen. Die Eigenschaftsanalyse wird im Kontext einer digitalen Speicherumgebung durchgeführt und ist maßgeblich für die Anforderungsanalyse und den sich anschließenden Systementwurf. Nur wer die relevanten Eigenschaften des zu bewahrenden Bezugsobjektes kennt, kann eine geeignete Aufbewahrungsumgebung konzipieren. Die Forderung nach einem Vorgehen, das auf den signifikanten Eigenschaften einer Web Site fußt und diese zur Definition einer geeigneten Sicherungsumgebung verwendet, lässt sich aus etablierten Standards extrahieren.<sup>52</sup> Die Struktur der Anforderungsanalyse wird über die Eigenschaften hinaus mit theoretisch fundierten und praktisch angewandten Prozessschritten zur Sicherung digitaler Güter angereichert und unter Berücksichtigung des Nutzens einer Web-Site-Biographie durchgeführt. Auf einer abstrakten Ebene lassen sich die Prozessschritte bereits hier in einen gefestigten und aus der Archivierung bekannten Dreiklang gliedern: *Erfassen*, *Bewahren* und *Bereitstellen*.<sup>53</sup> Der auf der Anforderungsanalyse basierende Systementwurf umfasst anschließend sowohl dynamische Elemente im Sinne von Prozessen als auch statische Elemente im Sinne von Dateiformaten, Datenbankstrukturen, Funktionen und Benutzeroberflächen. Diese bilden wiederum die Grundlage der prototypischen Realisierung. Der Prototyp selbst wird in der vorliegenden Arbeit abschließend beschrieben sowie zur Demonstration und Evaluation der Zielerreichung verwendet.

4. *Demonstration und Evaluation*: Die Demonstration und die Evaluation des Prototyps eines Web-Site-Biographie-Systems sind unmittelbar miteinander verbunden und folgen dem Entwurf bzw. der Entwicklung. Während die Demonstration auf das Präsentieren der Funktionen abzielt, stellt die Evaluation fest, in wie weit das forcierte IT-System zur Lösung des Problems und der übergeordneten Zielsetzung beiträgt.<sup>54</sup>

---

51 Vgl. Ludewig, Jochen; Lichter, Horst: *Software Engineering: Grundlagen, Menschen, Prozesse, Techniken*, 3., korrigierte Aufl., Heidelberg: dpunkt Verl. 2013, S. 353.

52 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): *Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644*, 1. Aufl, Berlin: Beuth 2013, S. 53 und Brown, Adrian: *Practical digital preservation: a how-to guide for organizations of any size*, a. a. O., S. 82 ff.

53 Vgl. Schiedermeier, Ute: *Wirtschaftsarchive am Beispiel des Siemens Historical Institute*, in: *AWV-Informationen, AWV-Informationen Special – Webarchivierung/2012*, S. 13.

54 Vgl. Pfeffers, Ken; Tuunanen, Tuure; Rothenberger, Marcus A.; Chatterjee, Samir: *A design science research methodology for information systems research*, a. a. O., S. 55 ff.

5. *Kommunikation*: Der letzte Schritt umfasst die Publikation bzw. Veröffentlichung der Forschungsergebnisse. Er hat zum Ziel, das neu geschaffene Artefakt mit der wissenschaftlichen Gemeinschaft und interessierten Praktikern zu teilen.

Mit der vorstehenden Erläuterung der Forschungsmethode ist der in Abbildung 2 dargestellte Aufbau der Arbeit verbunden.

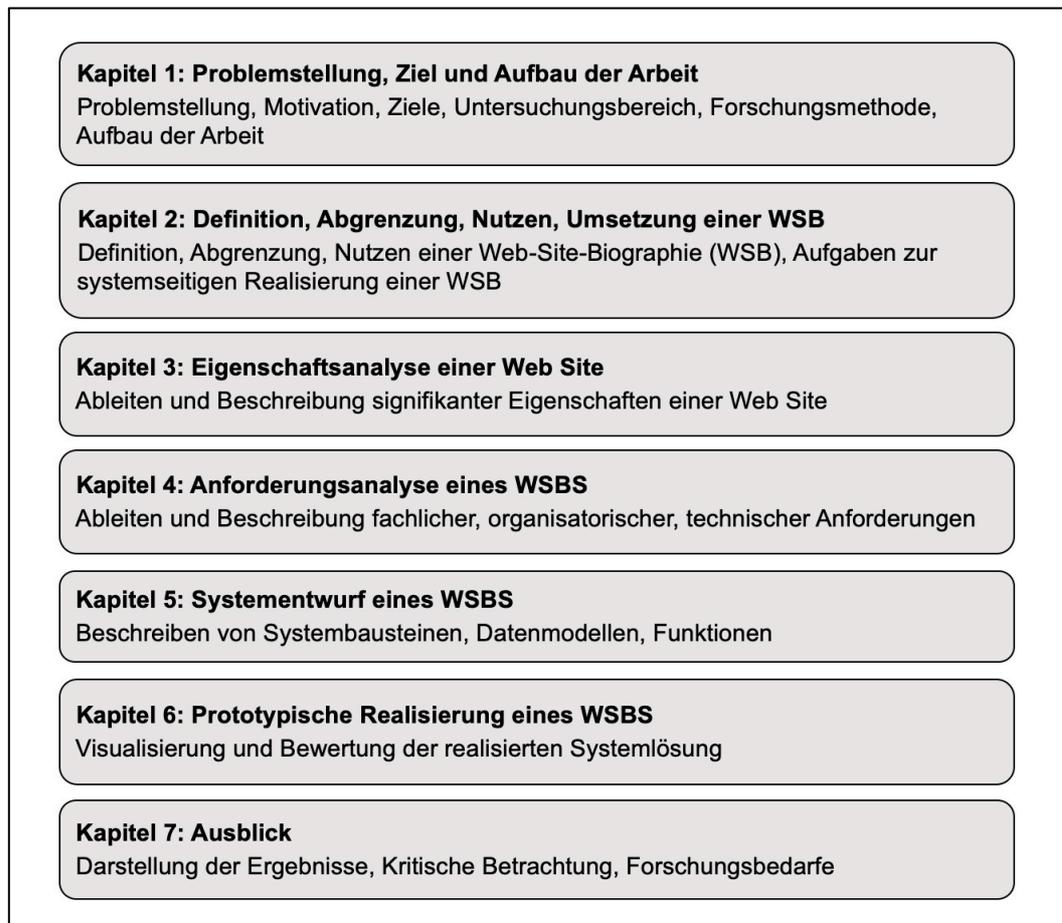


Abb. 2: Schematische Darstellung des Aufbaus der Arbeit

## 2 Definition, Abgrenzung, Nutzen, Umsetzung einer WSB

### 2.1 Systematisierung der Grundlagen

Im vorliegenden Kapitel der Arbeit wird zunächst der Begriff „Web-Site-Biographie“ definiert. Da die Wirtschaftsinformatik das Entwickeln nutzenstiftender Systeme in den Mittelpunkt stellt, wird darauf aufbauend der Nutzen einer Web-Site-Biographie für Unternehmen abgeleitet.<sup>55</sup> Im Anschluss an die Begriffs- und Nutzendefinition folgt eine Auseinandersetzung mit Aufgaben, die zur systemgestützten Dokumentation der Lebensgeschichte einer Web Site herangezogen werden können. Zur Gliederung der Aufgaben wird ein etablierter Dreiklang zum Dokumentieren digitaler Artefakte herangezogen: *Erfassen, Bewahren, Bereitstellen*.<sup>56</sup>

Für das *Erfassen* einer Web-Site-Biographie werden technische Instrumente erläutert, die z. B. von Suchmaschinenbetreibern verwendet werden, um Web Sites zu kopieren und zu sammeln. Aufgaben, die neben dem Erfassen auch das *Bewahren* und *Bereitstellen* einer Web-Site-Biographie unterstützen, werden mit Hilfe etablierter, normierender Rahmenwerke analysiert. Konkret sind an dieser Stelle der Standard 31644 des Deutschen Instituts für Normungen und die ISO-Norm 16363 zu nennen.<sup>57</sup> Das nachfolgende Kapitel beschäftigt sich zunächst mit der Definition des Begriffs „Web-Site-Biographie“.

### 2.2 Definition einer „Web-Site-Biographie“

Der Begriff „Web-Site-Biographie“ ist eine Komposition, die sich aus den Bestandteilen „Web Site“ und „Biographie“ zusammensetzt. Um eine vollständige Bestimmung des Begriffes zu erreichen, wird zunächst die „Web Site“ definiert und anschließend um die Definition des Begriffs „Biographie“ erweitert. Auf diese Weise soll eine stringente Bestimmung des zentralen Forschungsgegenstandes dieser Arbeit durchgeführt werden.

---

55 Vgl. Balzert, Helmut: Lehrbuch der Software-Technik, a. a. O., S. 35 f.

56 Vgl. unter anderem Gaus, Wilhelm: Berufe im Archiv-, Bibliotheks-, Informations- und Dokumentationswesen: ein Wegweiser zur Ausbildung, Berlin: Springer 1986, S. 13.

57 Vgl. Brown, Adrian: Practical digital preservation: a how-to guide for organizations of any size, a. a. O., S. 85.

### *Begriffsdefinition einer Web Site*

Im Allgemeinen wird unter einer „Web Site“ die digitale Präsenz einer Organisation verstanden.<sup>58</sup> Aus einer technischen Perspektive betrachtet, ist diese digitale Präsenz durch einen sach-logischen Verbund von HTML-Dateien charakterisiert, die auf Basis des Internet-Dienstes „World Wide Web (WWW oder Web)“ genutzt werden. Die einzelnen HTML-Dateien – im Folgenden auch „Web-Seiten“ genannt – werden über das Computer-Netzwerk „Internet“ zur Verfügung gestellt. Zur Codierung der Web-Seiten wird eine Beschreibungssprache verwendet, die sogenannte „Hypertext Markup Language (kurz: HTML)“. Das Hypertext-Transfer-Protokoll (kurz: „HTTP“) ermöglicht die Übertragung der codierten Inhalte auf Basis des Internet-Protokolls „TCP/IP“. Ein Web Browser macht die codierten Inhalte nutzbar.<sup>59</sup>

Alle Web-Seiten einer Web Site sind über eindeutige Adressen erreichbar. Eine eindeutige Adresse wird „Uniform Resource Locator (URL)“ genannt. Durch die Verwendung eines URL können Nutzer einzelne Web-Seiten ansteuern und anzeigen lassen.<sup>60</sup> Typischerweise fungiert dabei eine Web-Seite als Startseite der kompletten Web Site. Ausgehend von der Startseite können (alle) weitere(n) Web-Seiten der Web Site erschlossen werden. Die Navigationsgrundlage bilden „Hyperlinks (kurz: Links)“, die als Querverweise zwischen den Web-Seiten eingebaut werden.

Wurden in den Anfängen des WWW überwiegend statische Inhalte auf einer Web Site publiziert, ist das Web heute längst durch Dynamik geprägt. Web Sites werden durch Skriptsprachen erweitert, um interaktive Navigationselemente und andere reaktive Inhalte zu erzeugen. Programmiersprachen wie „PHP“, „Python“ oder „Java“ sind auf den Servern installiert, die zur Bereitstellung der Web Site eingesetzt werden. Algorithmen laden Inhalte aus Datenbanken und ermöglichen Entwicklern das Implementieren komplexer Anwendungen, die über das HTTP abgerufen werden. Neben der ursprünglichen Verwendung als statisches und meist lineares Informationssystem lassen sich heute nahezu sämtliche Aktivitäten entlang der primären und sekundären Wertschöpfung durch Internet-Technologie und das darauf aufbauende WWW realisieren.<sup>61</sup> Unternehmen verwenden die eigene Web Site nicht selten als Plattform, um auf einer einheitlichen Basis

---

58 Vgl. Springer Gabler Verlag (Hrsg.): Stichwort: Website, Online im Internet: <http://wirtschaftslexikon.gabler.de/Archiv/75916/website-v12.html>, 11.07.2017. und Schwickert, Axel C.: Web Site Engineering: Ökonomische Analyse und Entwicklungssystematik für eBusiness-Präsenzen, Stuttgart: Teubner 2001, S. 100.

59 Vgl. Schwickert, Axel C.: Web Site Engineering: Ökonomische Analyse und Entwicklungssystematik für eBusiness-Präsenzen, a. a. O., S. 100.

60 Vgl. Meinel, Christoph; Sack, Harald: WWW – Kommunikation, Internetworking, Web-Technologien, Berlin: Springer 2013, S. 12 ff.

61 Vgl. Meier, Andreas; Stormer, Henrik: eBusiness & eCommerce: Management der digitalen Wertschöpfungskette, 3. Aufl, Berlin: Springer Gabler 2012, S. 20.

von Funktionen zu kommunizieren. In diesem Zusammenhang ist das Individualisieren von Benutzeroberflächen längst keine Seltenheit mehr. Durch das Auswerten von Sitzungsdaten lassen sich problemlos personalisierte Inhalte erstellen.<sup>62</sup>

Von diesen längst etablierten Möglichkeiten der Personalisierung über komplexe dreidimensionale Spiele bis hin zu interaktiven Office-Lösungen und virtuellen Realitäten sind den Anwendungsmöglichkeiten einer Web Site kaum Grenzen gesetzt.<sup>63</sup> Die Web Site wird im digitalen Wirtschaftsgefüge daher längst zur Abwicklung unterschiedlichster Aktivitäten innerhalb einer Organisation und zwischen Organisationen eingesetzt. Nicht nur Kunden eines Unternehmens nutzen die Web Site, sondern auch Geschäftspartner und Mitarbeiter. Typischerweise stehen diesen drei Akteuren unterschiedliche Segmente zur Verfügung, um Inhalte abzurufen. Es wird im Regelfall zwischen den folgenden drei Segmenten unterschieden: Einem öffentlichen Bereich für Kunden („Internet“), einem geschlossenen Bereich für die bestehenden Geschäftspartner („Extranet“) und einem internen Bereich für die Mitarbeiter des Unternehmens („Intranet“).<sup>64</sup>

Der öffentliche Bereich einer Web Site, das Segment „Internet“, ist typischerweise für alle Nutzer des WWW verfügbar und dient unter anderem zur Anbahnung von geschäftlichen Transaktionen. Im öffentlichen Segment werden u. a. Produkte angeboten oder Dienstleistungen beworben. Zudem lassen sich Mehrwertdienste wie Nachrichten oder Suchfunktionen bereitstellen.<sup>65</sup> Das Extranet einer Web Site wird dagegen für Geschäftspartner, Mitglieder bzw. Kunden der Organisation vorgehalten und ist im Regelfall durch ein individuelles Benutzerkonto charakterisiert. Das Benutzerkonto kann nur nach einer erfolgreichen Authentifizierung (z. B. durch die Eingabe der korrekten Benutzername-Passwort-Kombination) abgerufen und verwendet werden. Es ist nicht für die Öffentlichkeit bestimmt und häufig an die Aktivitäten des Nutzers angepasst.

Der Intranet-Bereich einer Web Site ist ein ebenso abgeschlossenes Segment, steht aber ausschließlich den Mitarbeitern einer Organisation zur Verfügung.<sup>66</sup> Im Intranet werden interne Prozesse unterstützt und abgewickelt. Typische Aufgaben, die im Rahmen des

---

62 Vgl. Kappel, Gerti; Pröll, Birgit; Reich, Siegfried; Retschitzegger, Werner: Web Engineering – Die Disziplin zur systematischen Entwicklung von Web-Anwendungen, a. a. O., S. 7.

63 Vgl. Kappel, Gerti; Pröll, Birgit; Reich, Siegfried; Retschitzegger, Werner: Web Engineering – Die Disziplin zur systematischen Entwicklung von Web-Anwendungen, a. a. O., S. 7. und Krause, Dirk: VR im Web: Virtual-Reality-Web-Apps mit A-Frame programmieren, in: c't magazin für computer und technik, 16/2017, S. 116 ff.

64 Vgl. Schwickert, Axel C.: Web Site Engineering: Ökonomische Analyse und Entwicklungssystematik für eBusiness-Präsenzen, a. a. O., S. 93 ff.

65 Vgl. Bächle, Michael; Lehmann, Frank R.: E-Business: Grundlagen elektronischer Geschäftsprozesse im Web 2.0, München: Oldenbourg 2010, S. 4 f.

66 Vgl. Balzert, Heide: Webdesign & Web-Ergonomie: Websites professionell gestalten, Herdecke: W3L 2004, S. 13.

Web Content Managements (WCM) durchgeführt werden, sind die Entwicklung von Design-Vorlagen, das Erstellen und Verändern der Inhalte, die Definition von Arbeitsschritten oder die Administration von Berechtigungen für Benutzer oder Benutzergruppen. Der Intranet-Bereich bleibt der Öffentlichkeit und organisationsfremden Nutzen der Web Site verborgen.<sup>67</sup>

Aufbauend auf der obenstehenden Definition einer Web Site wird im anschließenden Abschnitt eine Begriffsbestimmung des zentralen Gestaltungsziels dieser Arbeit durchgeführt. Zunächst wird dafür der Begriff „Biographie“ alleinstehend beschrieben und darauf aufbauend die „Web-Site-Biographie“ definiert.

### *Begriffsdefinition einer Web-Site-Biographie*

Der Begriff „Biographie“ spielt die zweite wesentliche Rolle bei der Definition einer „Web-Site-Biographie“. In seiner ursprünglichen Form beschreibt die „Biographie“ eine retrospektive Darstellung eines Bezugsobjektes, die über einen definierten Zeitraum berichtet. Der Fokus liegt dabei im Regelfall auf einer Darstellung, die das Leben und Wirken eines Menschen präsentiert. Der inhaltliche Schwerpunkt und das Ziel der Biographie hängen von der fachlichen Domäne ab, die dieser retrospektiven Darstellung zugeordnet wird. Aus einer zeitlichen Perspektive betrachtet, kann bei der Definition einer „Biographie“ differenziert werden: Zum einen bestehen Definitionen für eine Biographie, die den kompletten Existenzzeitraum des Bezugsobjektes umfassen, zum anderen sind auch Definitionen vorhanden, die keinen fest definierten Zeitraum vorschreiben und sowohl theoretisch als auch praktisch eine Eingrenzung auf bestimmte Abschnitte tolerieren.<sup>68</sup> In der vorliegenden Untersuchung soll diese Eingrenzung nicht erfolgen. Die „Biographie“ wird als eine Darstellung verstanden, die in ihrer endgültigen Form über den kompletten Existenzzeitraum eines Bezugsobjektes (hier: Web Site) berichtet und keine Eingrenzung auf ein Zeitfenster innerhalb des Existenzzeitraums vornimmt.

Zusammengefasst kann eine „Web-Site-Biographie“ also bereits an dieser Stelle als retrospektive Darstellung bestimmt werden, die den kompletten Existenzzeitraum einer Web Site umfasst und dokumentiert. Die „Existenz“ einer Web Site wird dabei im Sinne des Vorhandenseins verstanden. Die Web Site einer Organisation existiert, solange Webseiten, Link-Strukturen, Inhalte, Gestaltungselemente und Segmente unter einer organisationsspezifischen Adresse im WWW vorhanden sind und von der jeweiligen Zielgruppe (z. B. Öffentlichkeit, Kunden, Geschäftspartner, Mitarbeiter) abgerufen und verwendet werden können. Der Zeitraum ist dabei nicht an eine bestehende Infrastruktur

---

67 Vgl. Krüger, Jörg Dennis; Kopp, Matthias: Web-Content managen: professioneller Einsatz von Content-Management-Systemen, a. a. O., S. 158.

68 Vgl. Momigliano, Arnaldo: The development of Greek biography, Expanded ed., 1st Harvard University Press pbk. ed, Cambridge, Mass: Harvard University Press 1993, S. 11.

gekoppelt, die den Betrieb des technischen Konstrukts auszeichnet. Vielmehr soll der Wechsel von Betriebselementen als Meilensteine in einer Web-Site-Biographie erfasst und für die Nachwelt konserviert werden. Im Gegensatz zu klassischen Archiven kann die Web-Site-Biographie nicht nur Inhalte originalgetreu dokumentieren, sondern auch Kontext-Informationen über die betriebspezifischen Gegebenheiten im „Leben“ der Web Site präsentieren.<sup>69</sup>

Die Verwendung des Begriffs „Leben“ ist grundsätzlich bzw. unmittelbar mit der Verwendung des Begriffs „Biographie“ verbunden und bedarf an dieser Stelle einer weiteren Ausführung. Berücksichtigt man die Wortherkunft der „Biographie“, setzt sich diese aus dem griechischen Wort „*Bios*“ für „Leben“ und „*gráphein*“ für „zeichnen“, „schreiben“ zusammen. Wird die Web Site also als Bezugsobjekt einer Biographie – und damit der Beschreibung eines Lebens – herangezogen, scheint der Terminus „Web-Site-Biographie“ zwei in sich verschiedene Konzepte zu kombinieren.<sup>70</sup> Die Web Site kann zwar als ein sozio-technisches System verstanden werden und neben Aktivitäten, Technologien und Verantwortlichkeiten auch den Menschen als Systemkomponente umfassen, im Kern ist die Web Site jedoch ein technisches Konstrukt.<sup>71</sup> Es muss daher darauf hingewiesen werden, dass der Begriff „Web-Site-Biographie (kurz: WSB)“ den Charakter einer *Metapher* aufweist und in diesem Sinne zu verstehen ist.<sup>72</sup>

Trotz seiner metaphorischen Verwendung beschreibt der Begriff „Web-Site-Biographie“ das Gestaltungsziel der vorliegenden Untersuchung in geeigneter Art und Weise. Im folgenden Kapitel soll diese Aussage verdeutlicht werden.

---

69 Vgl. Steinke, Tobias: Herausforderungen bei der Langzeitverfügbarkeit von Webarchiven, in: AWW-Informationen, AWW-Informationen Special – Webarchivierung/2012, S. 34 f.

70 Vgl. Goschler, Juliana: Metaphern für das Gehirn: eine kognitiv-linguistische Untersuchung, Berlin: Frank & Timme 2008, S. 48.

71 Vgl. Ferstl, Otto K.; Sinz, Elmar J.: Grundlagen der Wirtschaftsinformatik, 7., aktualisierte Aufl, München: Oldenbourg 2013, S. 66.

72 Vgl. vor allem die Eigenschaft der Übertragung bei Goschler, Juliana: Metaphern für das Gehirn: eine kognitiv-linguistische Untersuchung, a. a. O., S. 48.

### 2.3 Abgrenzung einer Web-Site-Biographie

Das Ziel der WSB ist das Erlebbarmachen der digitalen Vergangenheit einer Organisation.<sup>73</sup> Als Reproduktion verstanden, sammelt die WSB Zustände und Veränderungen einer Web Site, ordnet diese in einer Speicherumgebung und erzeugt ein nutzbares und interaktives Faksimile der Vergangenheit. Verwandte Konzepte, die sich aus einer geschichtswissenschaftlichen Motivation mit der Vergangenheit einer Web Site auseinandersetzen, lassen sich in der Literatur unter dem Begriff der „Web-Site-Historiographie“ finden.<sup>74</sup> Diese Form der historischen Interpretation soll im folgenden Abschnitt von einer WSB abgegrenzt werden. Dafür wird der Begriff erläutert und anschließend abgrenzende Merkmale dargelegt.

Die Bedeutung des WWW ist für unsere Gesellschaft unbestritten. Es ist daher nicht verwunderlich, dass sich neben technologiegeprägten Domänen der Wissenschaft auch Historiker mit diesem Informationssystem auseinandersetzen. Unlängst sind Werke veröffentlicht worden, die sich aus historischen Interessen mit dem WWW beschäftigen. Die Web-Site-Historiographie beschreibt dabei einen Prozess, der sich explizit auf die Gestaltung einer historischen Darstellung bezieht, die eine Web Site als Gegenstand hat. Bis heute stellt diese Form der Geschichtsschreibung jedoch eine nicht weit verbreitete Ausprägung des historischen Bewusstseins dar und ist exklusiver Bestandteil geschichtswissenschaftlicher Domänen.<sup>75</sup>

Das übergeordnete Ziel einer Geschichtsschreibung besteht darin, vergangene Ereignisse zu deuten, Wirkungszusammenhänge herzustellen und diese in einer narrativen Form festzuhalten.<sup>76</sup> Dabei fokussiert eine Web-Site-Historiographie Ereignisse und Wirkungszusammenhänge, die mit einer Web Site verbunden sind. Klassischerweise stellt ein Historiker dabei die Fragen nach dem „Wer?“, „Was?“, „Wann?“, „Warum?“ und „Wie?“, um Beteiligte festzustellen, Ereignisse zu definieren, chronologische Anordnungen vorzunehmen, Kausalitäten zu erkennen, den Einfluss auf die Gegenwart zu beschreiben oder eine mögliche Zukunft zu antizipieren.

---

73 Vgl. Rogers, Richard; Brügger, Niels (Hrsg.); Milligan, Ian (Hrsg.): *Periodizing Web Archiving: Biographical, Event-Based, National and Autobiographical Traditions*, in: *The SAGE Handbook of Web History*, Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne: SAGE Publishing 2018, S. 44 f.

74 Vgl. Brügger, Niels: *Web historiography and Internet Studies: Challenges and perspectives*, a. a. O., S. 3 ff.

75 Vgl. Brügger, Niels: *Web History, an Emerging Field of Study*, a. a. O., S. 5. Oder vgl. Brügger, Niels: *Web historiography and Internet Studies: Challenges and perspectives*, a. a. O., S. 3.

76 Vgl. Metzger, Franziska: *Geschichtsschreibung und Geschichtsdenken im 19. und 20. Jahrhundert*, in: *Historische Zeitschrift*, Band 296, Heft 3/2013, S. 5 f.

Grundsätzlich setzen die vorgenannten W-Fragen eine Summe von Aktivitäten voraus, die bei der Erstellung unterschiedlichster historischer Darstellungen eine Anwendung finden können.<sup>77</sup> Da die Biographie auch als eine Form der historischen Darstellung klassifiziert werden kann, ist dieses Vorgehen prinzipiell auf die Erstellung einer WSB übertragbar. Zudem wird die historische Grundfunktion der Belehrung und Unterhaltung nicht nur im Rahmen einer Web-Site-Historiographie, sondern auch im Rahmen der WSB berücksichtigt.<sup>78</sup> Das Vorgehen bei der Interpretation der Vergangenheit und die Grundfunktionen der Belehrung und Unterhaltung sollen daher nicht als abgrenzende Merkmale gesehen werden. Vielmehr markieren der Zeitraum und die angestrebte Realisierungsform der WSB die abgrenzenden Eigenschaften.

Bei einer Web-Site-Historiographie handelt es sich im Regelfall um eine retrospektive Zusammenfassung der Vergangenheit in Schriftform.<sup>79</sup> Diese Zusammenfassung wird als eine spezialisierte Auseinandersetzung mit Ereignissen, Objekten bzw. Personen der Vergangenheit verstanden. Das zentrale Thema, der Umfang sowie die berücksichtigten Personen, Objekte, Ereignisse und Fragestellungen können dabei durch den Autor, Herausgeber oder andere Einflussfaktoren bestimmt werden. Es existiert ein notwendiger Spielraum für die Ausgestaltung der Inhalte. Dieser Spielraum umfasst auch den betrachteten Zeitraum. Die Web-Site-Historiographie kann einen bestimmten Zeitpunkt, einen Zeitabschnitt, bestimmte Ereignisse oder auch ausgewählte Personen im Kontext der Web Site betrachten. Nicht zwingend wird eine Geschichte unter Berücksichtigung aller Inhalte, Kontexte und Wirkungszusammenhänge forciert.<sup>80</sup> Das Konzept einer WSB soll dabei helfen, diese Lücke zu schließen.

Die WSB ist daher in der vorliegenden Arbeit als digitale Dokumentation definiert, die den kompletten Existenzzeitraum einer Web Site berücksichtigt und möglichst sämtliche Zustände sowie Veränderungen abbildet, erläutert und *im WWW erlebbar macht*. Der Fokus liegt auf der Erhaltung und Präsentation der „Lebensgeschichte“ ohne Eingrenzung auf eine Menge von Inhalten, Gestaltungselementen oder betriebsspezifischen Gegebenheiten. Die WSB kann aus einer zeitlichen Perspektive also erst dann abgeschlossen werden, wenn die Web Site unter ihrem organisationsspezifischen Adressraum nicht mehr im WWW abrufbar ist.

---

77 Vgl. Brügger, Niels: Web History, an Emerging Field of Study, a. a. O., S. 5.

78 Vgl. von Zimmermann, Christian: Geschichte oder Biographie: Leopold Rankes Porträts Papst Pauls III. und Wallensteins, in: Die Biographie – Beiträge zu ihrer Geschichte, Hrsg.: von Hemecker, Wilhelm, Berlin, New York: De Gruyter 2009, S. 76.

79 Vgl. Metzger, Franziska: Geschichtsschreibung und Geschichtsdenken im 19. und 20. Jahrhundert, a. a. O., S. 5 f.

80 Vgl. Brügger, Niels: Web historiography and Internet Studies: Challenges and perspectives, a. a. O., S. 5 f.

Damit ein Rückblick im Sinne einer WSB erzeugt werden kann, ist ein qualifiziertes IT-System notwendig, das einen Bestand an Zuständen und Veränderungen nutzbar und erlebbar macht.<sup>81</sup> Das systematische Erfassen, Bewahren und Bereitstellen der Zustände einer Web Site wird dabei als Kernaufgabe des IT-Systems gesehen.<sup>82</sup> Möglichkeiten zur Realisierung dieser Kernaufgaben werden in den folgenden Kapiteln analysiert. Vorab beschäftigt sich Kapitel 2.4 zunächst mit dem Nutzen einer WSB für Unternehmen.

## 2.4 Zum Nutzen einer Web-Site-Biographie

Die Disziplin der Wirtschaftsinformatik stellt sich nicht nur der Frage, wie technische Lösungen für Unternehmen bzw. Organisationen skizziert und realisiert werden, sondern auch wie diese System-Lösungen nutzenstiftend verwendet werden können. Dabei ist in den meisten Fällen ein Balance-Akt zwischen möglicher technischer Realisierung und rein wirtschaftlicher Betrachtung zu absolvieren, um im Ergebnis ein geeignetes IT-System für die jeweilige Organisation zu entwickeln. Damit die Perspektive des Nutzens in der vorliegenden Arbeit nicht unberücksichtigt bleibt, wird der mögliche Nutzen einer WSB in den folgenden Abschnitten genannt und erläutert.<sup>83</sup>

Stellt man die Frage nach dem Nutzen einer WSB für Organisationen, lassen sich zwei wesentliche Verwendungsrichtungen feststellen, aus denen ein potentieller Nutzen abgeleitet werden kann. Zum einen lässt sich der Nutzen einer WSB durch die (Ab)Sicherung digitaler Bestände gegen Datenverlust und zur Wahrung der Nachweispflicht im WWW skizzieren. Zum anderen kann der Nutzen einer WSB durch die (Selbst)Darstellung der digitalen Vergangenheit für Kunden, Mitarbeiter oder Investoren beschrieben werden.<sup>84</sup>

### *Die WSB als (Ab)Sicherung*

Das langfristige Aufbewahren von Dokumenten wird innerhalb eines Unternehmens unter anderem durch Motivationen vorangetrieben, die aus Regelungen der Legislative oder Maßnahmen der Qualitätssicherung erwachsen. Klassische Treiber sind vor allem

---

81 Vgl. Brügger, Niels: When the present web is later the past: Web historiography, digital history, and Internet Studies, a. a. O., S. 105 oder Brügger, Niels: Web historiography and Internet Studies: Challenges and perspectives, a. a. O., S. 5 ff. und von Zimmermann, Christian: Geschichte oder Biographie: Leopold Rankes Porträts Papst Pauls III. und Wallensteins, a. a. O., S. 76.

82 Vgl. Gaus, Wilhelm: Berufe im Archiv-, Bibliotheks-, Informations- und Dokumentationswesen: ein Wegweiser zur Ausbildung, a. a. O., S. 13.

83 Vgl. dazu Mertens, Peter; Bodendorf, Freimut; König, Wolfgang; Schumann, Matthias; Hess, Thomas; Buxmann, Peter: Grundzüge der Wirtschaftsinformatik, 12., grundlegend überarbeitete Auflage, Berlin: Springer Gabler 2017, S. 2.

84 Vgl. Jochemich, Hanno: „Fehler HTTP 404 – Seite nicht gefunden“ Oder: Zum Stand der Archivierung von Webinhalten in den öffentlich-rechtlichen Rundfunkanstalten, a. a. O., S. 422 ff.

rechtliche Vorgaben, die Unternehmen zum Sichern geschäftsrelevanter Dokumente bewegen, um die Nachweispflicht im Rahmen juristischer Verfahren zu wahren.<sup>85</sup> Auch im Zusammenhang mit der Sicherung digitaler Dokumente hat z. B. die Neuregelung der Finanzmärkte im Jahr 2016 Einfluss auf die Bewahrung von im WWW publizierten Informationen.<sup>86</sup> Durch diese Neuregulierung sind Finanzinstitute im europäischen Rechtsraum verpflichtet, Informationen zu sichern, die im Rahmen von Verkaufsprozessen an Kunden weitergegeben werden und dies unabhängig vom zugrundeliegenden Medium. Das bedeutet, auch digitale Informationen, die beispielsweise auf der Web Site einer Bank publiziert werden, müssen zur Gewährleistung einer Nachweispflicht aus Unternehmensinteressen aufbewahrt und unter Umständen reproduziert werden. Das Interesse an geeigneten Sicherungsumgebungen, die sämtliche Kundenkommunikation im WWW rechtssicher dokumentieren, ist im Regelfall durch die Geschäftsführung begründet und soll die Geschäftsfähigkeit eines Unternehmens bzw. einer Organisation gewährleisten. Relevante Zustände einer Web Site müssen zum richtigen Zeitpunkt erfasst und zukunftsicher bewahrt werden. Die Notwendigkeit einer Bedienoberfläche, die jeden Zustand im Sinne einer lehrreichen und nostalgischen Dokumentation präsentiert und erlebbar macht, existiert dabei aber nicht. Vielmehr müssen Integrität und Authentizität der gesicherten Daten garantiert und vergangene Inhalte verlässlich reproduziert werden.<sup>87</sup>

Die (Ab)Sicherungsfunktion einer WSB kann einen vergleichbaren Nutzen umschreiben. Im Sinne eines auswertbaren Datenbestands können Zustände der Web Site dokumentiert und für die Organisation aufbewahrt werden. Da sich die technischen Gegebenheiten im Online-Bereich kontinuierlich weiterentwickeln, ist eine zentrale Herausforderung, Inhalte der Web Site auf unbestimmte Zeit in originalgetreuen Abbildern zu erhalten.<sup>88</sup>

### *Die WSB als (Selbst)Darstellung der digitalen Vergangenheit*

Neben der (Ab)Sicherung von Zuständen zur Wahrung der Nachweispflicht, kann eine WSB die digitale Vergangenheit für Kunden, Mitarbeiter, Investoren, Eigentümer oder das Management erlebbar machen. Im Gegensatz zu klassischen Unternehmenshistorien, die Vergangenes vor allem auf Basis von analogen Artefakten ausarbeiten und in Büchern

---

85 Vgl. Schneider, Holger: Digitale Amnesie: Langzeitarchivierung digitaler Dokumente im betrieblichen Umfeld, a. a. O., S. 22 f.

86 Vgl. Artikel 16 Absatz 7 in DAS EUROPÄISCHE PARLAMENT UND DER RAT DER EUROPÄISCHEN UNION (Hrsg.): RICHTLINIE 2014/65/EU DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 15. Mai 2014 über Märkte für Finanzinstrumente sowie zur Änderung der Richtlinien 2002/92/EG und 2011/61/EU, Online im Internet: <http://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32014L0065&from=EN>, 01.12.2018.

87 Vgl. Schneider, Holger: Digitale Amnesie: Langzeitarchivierung digitaler Dokumente im betrieblichen Umfeld, a. a. O., S. 41.

88 Vgl. dazu z. B. Jochemich, Hanno: „Fehler HTTP 404 – Seite nicht gefunden“ Oder: Zum Stand der Archivierung von Webinhalten in den öffentlich-rechtlichen Rundfunkanstalten, a. a. O., S. 424.

oder Museen präsentieren, können die historischen Zustände einer Web Site direkt im WWW reproduziert, verknüpft und wiedergegeben werden. Es wird im Idealfall eine interaktive Wanderung durch die digitale Vergangenheit des Unternehmens bzw. der Organisation auf Basis aktueller Technologien möglich. Auch wenn die Präsentationsumgebung von der einer klassischen Unternehmens-/Organisationsgeschichte abweicht, ist der Nutzen einer entsprechenden Darstellungsform vergleichbar. Aus einer ökonomischen Perspektive betrachtet, können Unternehmen bzw. Organisationen seit jeher durch die Präsentation der Vergangenheit Identität stiften, Alleinstellungsmerkmale kommunizieren, Mitarbeitern und Kunden eine Orientierungsmöglichkeit bieten, Lehren aus der Vergangenheit ziehen, Korrekturen vornehmen oder neue Ideen auf Basis vergangener Zustände und Ereignisse entwickeln.<sup>89</sup> Dieser Mehrwert ist auf das Darstellen der Lebensgeschichte einer Web Site übertragbar.

Das WWW wird tagtäglich von einer Vielzahl an Nutzern verwendet. Nicht nur um aktiv Inhalte zu generieren, sondern auch um unterschiedlichste Informationen über Unternehmen, Produkte oder Dienstleistungen zu erfragen. Die Web Site fungiert dabei als eine zentrale Anlaufstelle und hat zweifelsohne einen erheblichen Einfluss auf die Wahrnehmung der Organisation.<sup>90</sup> Teilweise ist die Web Site sogar das einzig wahrgenommene Abbild einer Organisation. Wird dieser Umstand berücksichtigt, kann nicht nur die klassische Unternehmensgeschichte, sondern konsequenterweise auch die WSB als identitätsstiftendes Instrument verstanden und genutzt werden. Durch das Darstellen vergangener Zustände einer Web Site kann die Qualität vergangener Web-Präsenzen präsentiert und damit potentielle Alleinstellungsmerkmale im WWW reproduziert und kommuniziert werden, z. B. das abwechslungsreiche, professionelle Design der Web Site im Gegensatz zu den bis heute weniger professionell entwickelten Präsenzen der Konkurrenz.<sup>91</sup>

Weiterhin ist denkbar, dass die WSB eine Orientierungsfunktion übernimmt. Durch das Verknüpfen der Abbilder einer Web Site und das Anreichern mit Kontext-Informationen können sowohl Investoren als auch Mitarbeiter oder Kunden durch die Vergangenheit reisen und die Evolution im digitalen Raum beurteilen. Für Kapitalgeber oder das Management kann es von Bedeutung sein, rückblickend festzustellen, wie die bereitgestellten Gelder über die Jahre im WWW investiert wurden, wann fundamentale Änderungen an der Web Site stattgefunden und in welcher Art sich diese auf den Erfolg ausgewirkt

---

89 Vgl. Sammer, Petra: *Storytelling*, 1. Aufl., 3. korrigierter Nachdruck, Beijing: O'Reilly 2015, S. 15 oder Berghoff, Hartmut: *Moderne Unternehmensgeschichte: eine themen- und theorieorientierte Einführung*, a. a. O., S. 14 ff.

90 Vgl. Haumer, Florian; Donsbach, Wolfgang: *Der Wertschöpfungsbeitrag von Corporate Publishing: Effekte formaler und inhaltlicher Gestaltungsmerkmale von Kundenmagazinen*, Wiesbaden: Springer VS 2013, S. 78 f.

91 Abgeleitet aus dem ökonomischen Nutzen einer Unternehmensgeschichte: Vgl. Berghoff, Hartmut: *Moderne Unternehmensgeschichte: eine themen- und theorieorientierte Einführung*, a. a. O., S. 15 ff.

haben, denn längst ist die Existenz einer Web Site an ihren wirtschaftlichen Erfolg geknüpft.<sup>92</sup> In diesem Zusammenhang ist es auch denkbar, dass die Zustände und Entwicklungsprozesse nicht nur in zeitlichen Abfolgen verkettet werden, sondern auch themenorientiert dokumentiert und mit erläuternden Texten präsentiert werden.

In Ergänzung zu den genannten Nutzen kann die Menge an historischen Daten nicht nur zum Reproduzieren und Erlebarmachen der Vergangenheit eingesetzt, sondern auch zum Generieren neuer Erkenntnisse auf Basis von Datenanalysen verwendet werden. Vergleichbar mit Anwendungen, die aus großen, strukturierten oder semi-strukturierten Datenbeständen neues Wissen generieren, kann potentiell auch aus den Datenbeständen einer WSB neues Wissen gewonnen werden. In diesem Zusammenhang sind unterschiedlichste qualitative oder quantitative Analysen denkbar.<sup>93</sup> Voraussetzung ist, dass der Datenbestand in einer angemessenen Weise vorgehalten und über eine geeignete Schnittstelle ausgewertet werden kann.<sup>94</sup>

Der Fokus der vorliegenden Arbeit liegt auf der Konstruktion eines IT-Systems, das zur (Selbst)Darstellung der digitalen Vergangenheit verwendet wird und eine WSB erlebbar macht. Eine Beschreibung der Aufgaben, die ein solches „Web-Site-Biographie-System (kurz: WSBS)“ durchführen muss, folgt im nächsten Kapitel.

---

92 Vgl. Hassler, Marco: Web Analytics: Metriken auswerten, Besucherverhalten verstehen, Website optimieren, 3., aktualisierte und erw. Aufl, Heidelberg Hamburg: mitp, Verl.-Gruppe Hüthig, Jehle, Rehm 2012, S. 28.

93 Vgl. Cleve, Jürgen; Lämmel, Uwe: Data mining, 2. Auflage, Berlin Boston: De Gruyter Oldenbourg 2016, S. 38 f.

94 In Anlehnung an die Analyse internationaler Links auf Nachrichten-Web-Sites: Vgl. Crowls, Josh; Blank, Grant: International hyperlinks in online news media, in: The web as history: using web archives to understand the past and the present, Hrsg.: Brügger, Niels; Schroeder, Ralph, London: UCL Press 2017, S. 101 ff. oder Jones, Steve; Johnson, Camille: Web Use and Web Studies, in: Web Archiving, Hrsg.: Masanès, Julien, Berlin, Heidelberg: Springer-Verlag 2010, S. 86 f. oder Abegglen, Martin: Geschichtsforscher werten Archive mit maschinellem Lernen aus, Online im Internet: <https://www.heise.de/newsticker/meldung/Geschichtsforscher-werten-Archive-mit-maschinellm-Lernen-aus-3940801.html>, 29.01.2018.

## 2.5 Aufgaben zur systembasierten Umsetzung einer WSB

### 2.5.1 Systematisierung

Die Aufgaben eines Web-Site-Biographie-Systems (kurz: WSBS) lassen sich aus dem etablierten Aufgabendreiklang des Archiv-, Bibliotheks- und Dokumentationswesens ableiten. Der Dreiklang ist auf Grund seines generischen Charakters für ein WSBS anwendbar und lässt sich durch das *Erfassen*, *Bewahren* und *Bereitstellen* digitaler Abbilder einer Web Site beschreiben.<sup>95</sup>

Die notwendigen Mechanismen zur Realisierung dieser Aufgaben unterscheiden sich von den klassischen Mechanismen zur Erstellung analoger Dokumentationen, Archive oder Bibliotheken. Analoge Objekte wie Bücher, Schriftstücke oder Bilder können im Regelfall ohne das Anfertigen von Kopien für die Nachwelt erfasst und gesichert werden. Inhalte, die im WWW publiziert werden, sind dagegen nicht ohne Weiteres in ihrer ursprünglichen Form und Umgebung nutzbar. Web Sites müssen aus dem zugrundeliegenden Informationssystem kopiert werden, bevor eine Sicherung für die Nachwelt möglich ist. Zudem lassen sich traditionelle Erhaltungsmethoden nicht unmittelbar auf digitale Formate übertragen. Gründe liegen unter anderem in der Beschaffenheit der digitalen Bezugsobjekte und in der Beschaffenheit der Datenträger, die zum physischen Bewahren der Informationsstrukturen verwendet werden.<sup>96</sup>

Etablierte Referenzmodelle wie das „Open Archiv Information System (OAIS)“ oder der „DCC Curation Lifecycle“ adressieren diese Problemstellung und skizzieren standardisierte Modelle zum Erhalten digitaler Artefakte.<sup>97</sup> Da diese Modelle einen generischen Charakter aufweisen, werden keine Spezifika zur Erfassung, Bewahrung oder Bereitstellung von Web Sites modelliert. Der Fokus dieser Referenzmodelle liegt auf allgemeingültigen Konzepten, die je nach Bezugsobjekt weiter spezifiziert werden müssen.

Ein angepasstes Modell zur Beschreibung der Aufgaben eines WSBS wird in der vorliegenden Arbeit aus den existierenden Modellen abgeleitet. Die Aufgaben eines WSBS

---

95 Vgl. Gaus, Wilhelm: Berufe im Archiv-, Bibliotheks-, Informations- und Dokumentationswesen: ein Wegweiser zur Ausbildung, a. a. O., S. 18 f.

96 Vgl. Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 62 f.

97 Vgl. Consultative Committee for Space Data Systems (Hrsg.): REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS), Online im Internet: <https://public.ccsds.org/pubs/650x0m2.pdf>, 12.08.2017. oder auch Harvey, D. R.; Gillian, Oliver: Digital curation, Second edition, Chicago: ALA Neal-Schuman, an imprint of the American Library Association 2016, S. 31 ff.

werden im Laufe der Arbeit durch Anforderungen konkretisiert, die sich aus den Eigenschaften des Bezugsobjektes „Web Site“ ergeben. Durch dieses Vorgehen soll im Ergebnis ein originärer Systementwurf entstehen.<sup>98</sup>

Abbildung 3 stellt die drei Aufgaben des WSBS dar und unterbaut sie mit den grundlegenden Systemmodulen: „Web Site analysieren“, „Abbilder generieren“, „Abbilder erschließen“, „Abbilder erhalten“ und „Abbilder bereitstellen“.<sup>99</sup>

Zur Konfiguration und Nutzung der Systemmodule (kurz: Module) wird eine Systemverwaltung integriert, die Bedienoberflächen bereitstellt.

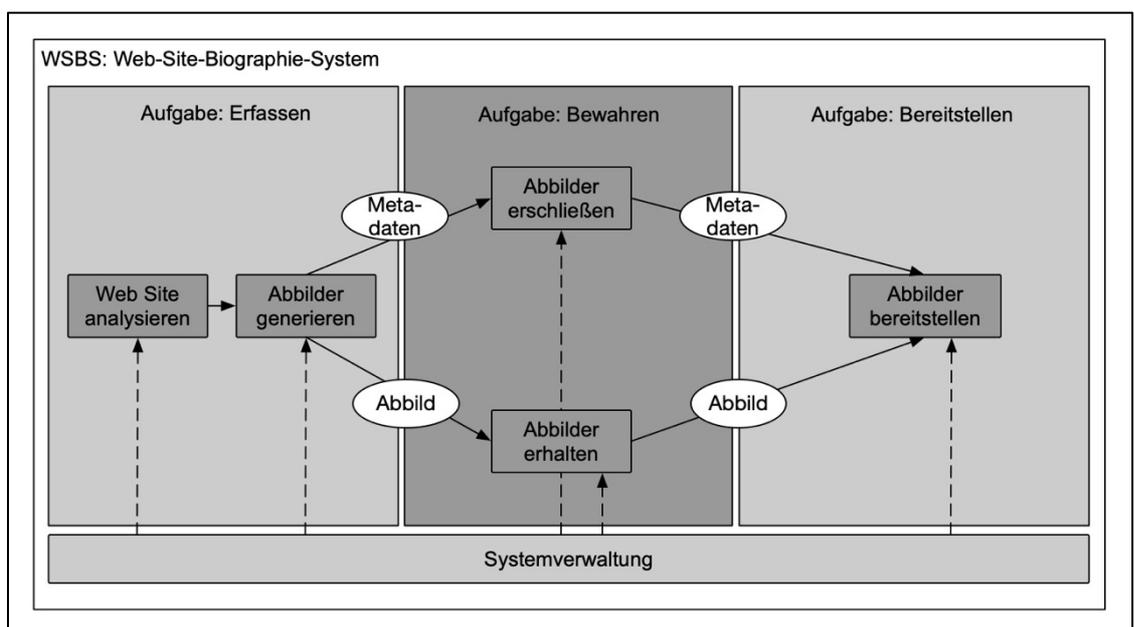


Abb. 3: Strukturmodell: Aufgaben und Module des WSBS

In den folgenden drei Kapiteln werden die zuvor beschriebenen Aufgaben und Module eines WSBS anhand bestehender IT-Werkzeuge (z. B. Web Crawler) und Konzepte (ISO-Normen oder DIN-Normen) konkretisiert.

98 Vgl. zu diesem Vorgehen auch Rosenthal, David; Robertson, Thomas; Lipkis, Tom; Reich, Vicky; Morabito, Seth: Requirements for Digital Preservation Systems – A Bottom-Up Approach, Online im Internet: <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>, 12.08.2017.

99 Vgl. Schimpf, Sabine: Das OAIS-Modell für die Langzeitarchivierung: Anwendung der ISO 14721 in Bibliotheken und Archiven, 1. Aufl, Berlin: Beuth 2014, S. 86 oder Consultative Committee for Space Data Systems (Hrsg.): REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS), a. a. O., 12.08.2017. oder Harvey, D. R.; Gillian, Oliver: Digital curation, a. a. O., S. 31 ff.

## 2.5.2 Erfassen

Das Erfassen einer WSB ist im WSBS durch die Module *Web Site analysieren* und *Abbilder generieren* charakterisiert. Mit Hilfe geeigneter Erfassungstechniken müssen die Zustände einer Web Site überwacht und auf Veränderungen überprüft werden. Sobald das WSBS einen bewahrungswürdigen Zustand feststellt, ist dieser in einem Abbild festzuhalten. Das Abbild wird mit geeigneten *Metadaten* angereichert und zur Bewahrung übergeben. Die Metadaten beinhalten im Regelfall technische, deskriptive, strukturelle und administrative Informationen zur Interpretation der Abbilder.<sup>100</sup> Neben Zugriffsrechten oder Informationen über Dateiformate können beispielsweise Kontext-Informationen zur Beschreibung der IT-Infrastruktur einer Web Site übermittelt und so die historische Nachvollziehbarkeit der Betriebsumgebung sichergestellt werden. Zur technischen Realisierung der Erfassung sind drei Verfahren bekannt: Es wird zwischen einer Client-seitigen, einer transaktionalen und einer Server-seitigen Erfassung differenziert.<sup>101</sup>

Bei der Client-seitigen Erfassung werden IT-Werkzeuge verwendet, die häufig unter dem Namen „Crawler“ subsumiert werden. Ein Crawler adaptiert das Verhalten eines Internet-Nutzers. Der Crawler öffnet eine Web Site und kopiert Inhalte und die Gestaltungselemente in ein externes Verzeichnis oder eine Datenbank. Typischerweise bewegt sich ein Crawler dabei von einer Web-Seite zur nächsten. Er indexiert die Links im HTML-Quellcode und wandert entlang dieser virtuellen Pfade, um eine Web Site möglichst vollständig zu erfassen.<sup>102</sup>

Die Anzahl der virtuellen Pfade, die ein Crawler abarbeiten muss, hängt wesentlich vom Anwendungsbereich ab. Grundsätzlich kann zwischen dedizierten und universellen Crawlern unterschieden werden. Dedizierte Crawler haben die Aufgabe, eine spezifische Web Site oder eine Menge von Web Sites zu einer spezifischen Themenstellung oder innerhalb einer spezifischen Netzwerk-Domäne (z. B. alle Web Sites der Second-Level-Domain „.de“) zu sichern. Universelle Crawler versuchen dagegen das komplette WWW zu erfassen. Typische Anwender dieser universellen Crawler sind Suchmaschinen wie *Google*, *Yahoo!* oder *Bing*.<sup>103</sup>

---

100 Vgl. Miller, Steven J.: *Metadata for digital collections: a how-to-do-it manual*, New York: Neal-Schuman Publishers 2011, S. 10 ff.

101 Vgl. Masanès, Julien: *Web Archiving: Issues and Methods*, in: *Web Archiving*, Hrsg.: Masanès, Julien, Berlin, Heidelberg: Springer-Verlag 2010, S. 22 ff.

102 Vgl. Pant, Gautam; Srinivasan, Padmini; Menczer, Filippo: *Crawling the Web*, in: *Web dynamics: adapting to change in content, size, topology and use ; with 29 tables*, Hrsg.: Levene, Mark; Poulouvassilis, Alexandra, Berlin: Springer 2004, S. 153. oder Alpar, Andre; Koczy, Markus; Metzen, Maik: *SEO - Strategie, Taktik und Technik: Online-Marketing mittels effektiver Suchmaschinenoptimierung*, Wiesbaden: Springer Gabler 2015, S. 47 f.

103 Vgl. Menczer, Filippo: *Web Crawling*, in: *Web Data Mining – Exploring Hyperlinks, Contents, and Usage Data*, Hrsg.: Liu, Bing, Berlin, New York: Springer-Verlag 2007, S. 273.

Für das Erfassen einer WSB ist die vergleichsweise einfache Installation und Pflege eines Crawlers von Vorteil.<sup>104</sup> Der erfolgreiche Betrieb eines Crawlers setzt keine tiefgreifende Integration in die IT-Infrastruktur einer Organisation voraus und kann unabhängig von den Betriebskomponenten einer Web Site erfolgen. Solange das Crawler-System die Web Site einer Organisation über das WWW abrufen kann, lassen sich Kopien einzelner oder aller Web-Seiten erzeugen. Erneuerungen des Server-Betriebssystems oder des Web-Servers zur Bereitstellung der Web Site haben typischerweise keine Auswirkungen auf die Installation und Betriebsfähigkeit des Crawlers.

Neben den genannten Vorteilen ist der Einsatz eines Crawlers aber auch mit Nachteilen verbunden. Da das WWW längst durch volatile und flüchtige Inhalte geprägt ist, sind die Erfassungsalgorithmen der Crawler durch Iterationen geprägt. Es reicht nicht aus, einmalig sämtliche HTML-Dateien bzw. Web-Seiten innerhalb eines spezifischen Adressraums zu erfassen. Vielmehr muss ein Crawler periodisch oder ereignisgesteuert die Erfassung wiederholen und die Web-Seiten innerhalb seines Einsatzgebietes abfragen. Nur so lassen sich die unterschiedlichen Zustände einer kompletten Web Site möglichst umfangreich in der dafür vorgesehenen Umgebung abbilden.<sup>105</sup>

In einigen Fällen ändern sich Inhalte aber unmittelbar und mehrmals, nachdem diese durch den Crawler gesichert wurden. Da die Erfassung typischerweise sequentiell abläuft, können Änderungen auf einer bereits gesicherten Web-Seite erst dann wieder erfasst werden, wenn die aktuelle Iteration beendet ist und ein neuer Durchlauf gestartet wird.<sup>106</sup> Als Konsequenz fehlen in der Speicherumgebung Inhalte, die sich zwischen Erfassungsprozessen mehrfach verändern.<sup>107</sup>

Ergänzend zur Volatilität stellen abgesicherte und personalisierte Inhalte eine weitere Herausforderung dar. Ohne Kenntnisse geeigneter Zugangsdaten bleiben abgesicherte Inhalte nicht nur der Öffentlichkeit, sondern auch den Crawlern verborgen. Zwar können einem Crawler Zugangsdaten übergeben und damit eine Kopie des abgesicherten Segmentes erzeugt werden, personalisierte Inhalte wie Produktvorschläge, die für jeden Nutzer individuell generiert werden, lassen sich jedoch nicht erfassen.<sup>108</sup>

---

104 Vgl. Fitch, K.: Web site archiving: an approach to recording every materially different response produced by a Website, Sanctuary Cove, Australia: The Ninth Australian World Wide Web Conference 2003, S. 5.

105 Vgl. Menczer, Filippo: Web Crawling, a. a. O., S. 273.

106 Vgl. Xavier, Roche: Copying Websites, in: Web Archiving, Hrsg.: Masanès, Julien, Berlin, Heidelberg: Springer-Verlag 2010, S. 95.

107 Vgl. Menczer, Filippo: Web Crawling, a. a. O., S. 274.

108 Vgl. Mitchell, Ryan E.: Web scraping with Python: collecting data from the modern web, 2. Auflage, Beijing, Boston, Farnham, Sebastopol, Tokyo: O'Reilly Media 2018, S. 158 ff.

Für historisch motivierte Initiativen, die mehr an einer Bestandsaufnahme interessiert sind als an der Dokumentation aller Nutzeraktivitäten, ist dieser Umstand jedoch nicht als nachteilig zu bewerten. Für Unternehmen, die z. B. durch den Gesetzgeber zum langfristigen Sichern personalisierter Inhalte angehalten sind, scheint der Einsatz eines Crawlers jedoch nicht zielführend. Das Sichern der Nutzeraktivitäten ist nicht realisierbar. Eine Alternative stellt das transaktionale Erfassungsverfahren dar.<sup>109</sup>

Das transaktionale Erfassungsverfahren filtert und kopiert alle Datenübertragungen, die zwischen Nutzern und der Web Site stattfinden. Sobald eine Anfrage von einem Web Server beantwortet wird, werden die Inhalte der Transaktion abgefangen, kopiert und weitergeleitet.<sup>110</sup> Die Funktionsweise ist vergleichbar mit einem Netzwerk-Proxy und beinhaltet das Potential, sämtliche Nutzeraktivitäten zu sichern.<sup>111</sup>

Für die Implementierung eines transaktionalen Erfassungsverfahrens ist eine technische Integration in die Bereitstellungssysteme einer Web Site notwendig. Der damit verbundene Konfigurations- und Pflegeaufwand ist vergleichsweise hoch. Darüber hinaus gilt es zu beachten, dass in einer transaktionsorientierten Speicherumgebung nur Abbilder aufbewahrt werden können, die tatsächlich von einem Nutzer angefragt und abgerufen wurden. Inhalte, die nicht durch Nutzer abgefragt werden, können auch nicht erfasst werden. Für historisch motivierte Initiativen, die typischerweise mehr an einer umfassenden Dokumentation der Web Site interessiert sind als an einzelnen Nutzeraktivitäten, sollte dieser Umstand unbedingt mitberücksichtigt werden.<sup>112</sup> Weiterhin ist das Nutzarmachen der Datenmengen als nachteilig zu bewerten. Die Inhalte einer Transaktion sind für eine historische Nutzarmachung innerhalb einer Speicherumgebung unter Umständen aufwendig nachzubereiten. Besonders das Filtern redundanter Abbilder ist mit teilweise erheblichen Ressourcenaufwendungen verbunden. Jedes Mal, wenn eine Web Site durch einen Nutzer aufgerufen wird, werden Abbilder der Transaktionen erzeugt. Unabhängig davon, ob Änderungen an den Inhalten stattgefunden haben oder nicht. Um das Problem redundanter Abbilder zu lösen, muss eine Programm-Komponente existieren, die identische Abbilder erkennt und eliminiert.<sup>113</sup>

---

109 Vgl. Fitch, K.: Web site archiving: an approach to recording every materially different response produced by a Website, a. a. O., S. 17 f.

110 Vgl. Masanès, Julien: Web Archiving: Issues and Methods, a. a. O., S. 26.

111 Vgl. Fitch, K.: Web site archiving: an approach to recording every materially different response produced by a Website, a. a. O., S. 2 oder Masanès, Julien: Web Archiving: Issues and Methods, a. a. O., S. 26.

112 Vgl. Masanès, Julien: Web Archiving: Issues and Methods, a. a. O., S. 26.

113 Vgl. Fitch, K.: Web site archiving: an approach to recording every materially different response produced by a Website, a. a. O., S. 2 ff. oder Masanès, Julien: Web Archiving: Issues and Methods, a. a. O., S. 26 f.

Neben der langfristigen Sicherung auf Basis eines transaktionalen Verfahrens soll abschließend die Server-seitige Erfassung erläutert werden.

Die Server-seitige Erfassung basiert auf IT-Werkzeugen, die durch einen direkten Zugriff auf die Datenbanken, Verzeichnisse und Programm-Codes einer Web Site charakterisiert sind. Durch den unmittelbaren Zugriff auf die Server einer Organisation kann beispielsweise die komplette Datenbank zum Betrieb einer Web Site kopiert werden. Auf diese Weise lassen sich auch Inhalte kopieren, die eigentlich durch Passwörter gesichert sind oder für ein Nutzer-Profil erzeugt werden.<sup>114</sup> Problematisch ist dabei, dass sämtliche Inhalte einer Web Site zunächst nur innerhalb der kopierten Datenbank vorliegen. Erst wenn Funktionen existieren, die auf die kopierte Datenbank zugreifen und daraus nutzbare Abbilder erzeugen, kann eine Dokumentation im Sinne einer WSB erfolgen. Ein solcher Reproduktionsmechanismus kann auf lange Frist kaum innerhalb einer Organisation implementiert werden, da der Aufwand in keinem angemessenen Verhältnis zum Nutzen der Systemlösung steht.

Bereits an dieser Stelle soll konstatiert werden, dass alleinstehend weder die Server-seitige noch die transaktionale Erfassung für das Erzeugen einer WSB sinnvoll erscheint.<sup>115</sup> Die transaktionale und Server-seitige Erfassung ermöglichen zwar theoretisch Dokumentationen sämtlicher Nutzeraktivitäten, setzen aber eine durchaus tiefgreifende Integration in die IT-Infrastruktur voraus. Der Crawler kann dagegen flexibel auf unterschiedlichste Infrastrukturen angepasst werden und ermöglicht das sequentielle bzw. ereignisgesteuerte Erfassen einer Web Site. Abgesicherte Inhalte können mit geeigneten Zugangsdaten kopiert werden.<sup>116</sup>

---

114 Vgl. Masanès, Julien: Web Archiving: Issues and Methods, a. a. O., S. 27 f.

115 Vgl. Masanès, Julien: Web Archiving: Issues and Methods, a. a. O., S. 28.

116 Vgl. Fitch, K.: Web site archiving: an approach to recording every materially different response produced by a Website, a. a. O., S. 3 ff.

Abbildung 4 hebt die Module zur Erfassung einer WSB abschließend durch eine Umrahmung im Strukturmodell des WSBS hervor.

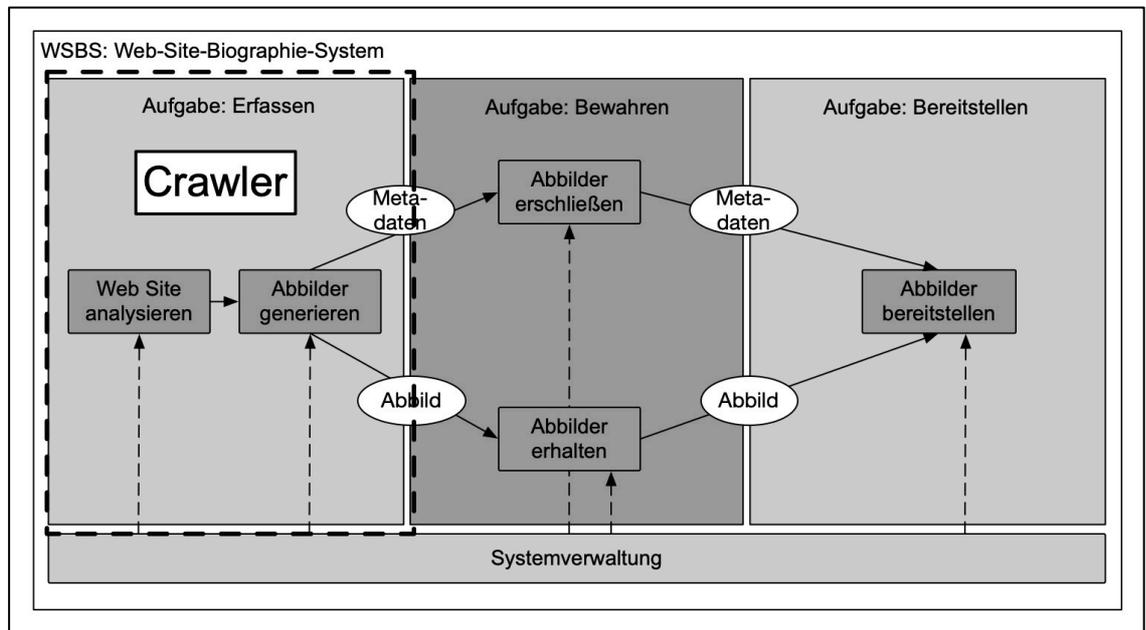


Abb. 4: Erfassungsmodule im Strukturmodell des WSBS

### 2.5.3 Bewahren

Das Erfassen von Abbildern ist im WSBS unmittelbar mit dem *Bewahren* verbunden. Eine wesentliche Herausforderung im Rahmen der Bewahrung ist die Erhaltung digitaler Datenbestände, denn traditionelle Methoden der Bestandserhaltung lassen sich nicht ohne Weiteres auf digitale Formate übertragen. Gründe liegen unter anderem in der Beschaffenheit der digitalen Objekte selbst und in der Beschaffenheit ihrer Datenträger.<sup>117</sup>

Im Kontrast zu analogen Datenträgern haben digitale Datenträger eine vergleichsweise geringe Haltbarkeit.<sup>118</sup> Zudem verwenden Unternehmen häufig proprietäre Software zum Erzeugen und Verarbeiten digitaler Datenbestände. Dieser Umstand bedingt, dass die Interpretation einzelner Daten häufig vom Einsatz einer speziellen Software abhängig ist.<sup>119</sup>

117 Vgl. Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 62 f.

118 Vgl. Schneider, Holger: Digitale Amnesie: Langzeitarchivierung digitaler Dokumente im betrieblichen Umfeld, a. a. O., S. 29 f. oder Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 4 ff.

119 Vgl. Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, Hamburg: Kovač 2013, S. 58 f.

Auch für Web Sites lassen sich in diesem Zusammenhang Anhaltspunkte finden, wie die langjährige Nutzung der Software „Flash“ zum Generieren von interaktiven Web-Anwendungen oder Spielen. Flash-basierte Inhalte lassen sich in einer Speicherumgebung nur dann reproduzieren, wenn die passende Flash-Installation zur Interpretation vorhanden ist. Ist die Software veraltet oder für Nutzer nicht mehr verfügbar, müssen die Inhalte in einer Emulationsumgebung präsentiert oder in andere Formate migriert werden. Ohne entsprechende Erhaltungsmaßnahmen ist der Inhalt nicht mehr nutzbar.<sup>120</sup>

Durch die vielfältigen Möglichkeiten digitale Artefakte zu produzieren, sind Forderungen nach geeigneten Lösungen zur Bewahrung dieser Artefakte akut.<sup>121</sup> Analysiert man die Fachliteratur, ist feststellbar, dass in diesem Zusammenhang etablierte Werke existieren, die für Organisationen normierende Hinweise zur Gestaltung digitaler Speicher- und Reproduktionsumgebungen bereitstellen.<sup>122</sup>

An dieser Stelle sind u. a. der Data Seal of Approval (DSA), die DIN-Norm 31644 und die ISO-Norm 16363 zu nennen. Die Relevanz dieser drei Rahmenwerke wird nicht zuletzt durch Bemühungen der Europäischen Union unterstrichen, diese Standards in einem dreistufigen Zertifizierungsverfahren zu kumulieren, dem sogenannten „European Framework for Audit and Certification (EFAC)“.<sup>123</sup> Das EFAC nimmt unmittelbar Bezug auf die Inhalte des DSA, der DIN-Norm 31644 sowie der ISO-Norm 16363 und kommuniziert eine schrittweise Zertifizierung auf Basis dieser Werke.<sup>124</sup>

Da es sich bei den zuvor genannten Rahmenwerken um generische Konzepte handelt, gehen diese nicht auf Forderungen ein, die durch die zu sichernden Objekte selbst (hier: Web Sites) entstehen. Gerade diese Forderungen sind jedoch entscheidend. Letztendlich bestimmen die Eigenschaften einer Web Site, welche Funktionen zum Bewahren integriert werden müssen.<sup>125</sup> Im Gegensatz zu einem einfachen Textdokument bedarf es z. B.

---

120 Vgl. Schneider, Holger: Digitale Amnesie: Langzeitarchivierung digitaler Dokumente im betrieblichen Umfeld, a. a. O., S. 119 ff. oder Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 52 ff.

121 Vgl. Budras, Corinna: Für immer verloren, in: Frankfurter Allgemeine Sonntagszeitung (F.A.S), 19/2018, S. 26.

122 Vgl. Brown, Adrian: Practical digital preservation: a how-to guide for organizations of any size, a. a. O., S. 82 ff.

123 Vgl. Schumann, Natascha: Einführung in die digitale Langzeitarchivierung, in: Langzeitarchivierung von Forschungsdaten: Standards und disziplinspezifische Lösungen, Hrsg.: Altenhöner, Reinhard; Oellers, Claudia, Berlin: Scivero Verl. 2012, S. 45.

124 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 30.

125 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 56 oder auch Rosenthal, David; Robertson, Thomas; Lipkis, Tom; Reich, Vicky; Morabito, Seth: Requirements for Digital Preservation Systems – A Bottom-Up Approach, a. a. O., 12.08.2017.



gesichert werden, die es Betreibern und Systemfunktionen ermöglichen, einzelne Abbilder zu finden und zu interpretieren. Die Metadaten können dabei in deskriptive, technische, administrative und strukturelle Metadaten gegliedert werden.

Deskriptive Metadaten beinhalten beschreibende Merkmale, die Systemnutzer zur historischen Interpretation der Zustände einer Web Site benötigen.<sup>129</sup> Technische, administrative und strukturelle Metadaten werden dagegen vor allem von Systemfunktionen benötigt, um die Reproduktion der Abbilder auf Basis von Programmkomponenten und Benutzeroberflächen zu ermöglichen. In technischen Metadaten werden z. B. Informationen über die Dateilänge und den Dateityp eines Abbildes gespeichert, administrative Metadaten führen dagegen Informationen über Zugriffsrechte einzelner Benutzer und strukturelle Metadaten stellen Verknüpfungen zwischen einzelnen Abbildern her.<sup>130</sup>

Zum Erhalten der Abbilder wird der binäre Zeichenstrom auf digitale Datenträger geschrieben. Geeignete Funktionen müssen bei Bedarf das Initiieren von Erhaltungsmaßnahmen ermöglichen. Beispielhaft ist hier das Migrieren veralteter Dateiformate zu nennen.<sup>131</sup> Um die Anzahl der Migrationen und damit auch das Risiko fehlerhafter Migrationsprozesse zu minimieren, müssen langfristig stabile bzw. interpretierbare Dateiformate gewählt werden. Der Quellcode eines Abbildes muss strukturiert gespeichert werden und auf unbestimmte Zeit reproduzierbar sein. Der Einsatz standardisierter und qualifizierter Dateiformate ist daher unbedingt notwendig.<sup>132</sup>

Grundsätzlich lassen sich in der Fachliteratur unterschiedliche Dateiformate feststellen, die zur Bewahrung bzw. Erhaltung von Abbildern einer Web Site empfohlen werden.<sup>133</sup> In der Praxis findet vor allem das „WebARChiv“-Format (kurz: WARC) Anwendung. Das WARC-Format ist ein speziell für die Langzeiterhaltung von Web Sites eingesetztes Dateiformat und wurde für die Bewahrung von Abbildern im Internet-Archiv entwickelt. Als international anerkannter Formatstandard wird das WARC-Format heute in einer

---

129 Vgl. z. B. Miller, Steven J.: Metadata for digital collections: a how-to-do-it manual, a. a. O., S. 30.

130 Vgl. Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 37.

131 Vgl. Ullrich, Dagmar: Bitstream Preservation, in: nestor Handbuch: eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 2.3, 2010, Hrsg.: Neuroth, Heike; Obwald, Achim; Scheffel, Regine; Strathmann, Stefan; Jehn, Mathias, 2., aktualisierte Auflage, Glückstadt: Verlag Werner Hülsbusch, Fachverlag für Medientechnik und -wirtschaft 2016, S. 165 f.

132 Vgl. Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 65.

133 Vgl. Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 60 f.

Vielzahl von Archivierungsinitiativen verwendet und ermöglicht eine einheitliche Konservierung von Web Sites.<sup>134</sup> Abbilder im WARC-Format können zudem in eine Vielzahl WARC-kompatibler Systemlösungen importiert und weiterverwendet werden. Die Wiederverwendbarkeit ist ein Aspekt der langfristigen Bestandserhaltung und muss auch bei der Realisierung eines WSBS berücksichtigt werden.<sup>135</sup>

#### 2.5.4 Bereitstellen

Das Bereitstellen einer WSB ist durch das Erleben sämtlicher Abbilder einer Web Site charakterisiert. Das Modul *Abbilder bereitstellen* macht die Abbilder für Nutzer im WWW verfügbar. Die einzelnen Abbilder können auf Basis ihrer Metadaten gesucht und in einer geeigneten Web-Oberfläche wiederhergestellt werden.

Abbildung 6 hebt den Aufgabenbereich und das Modul zur Bereitstellung der Abbilder durch eine gesonderte Umrahmung im Strukturmodell des WSBS hervor.

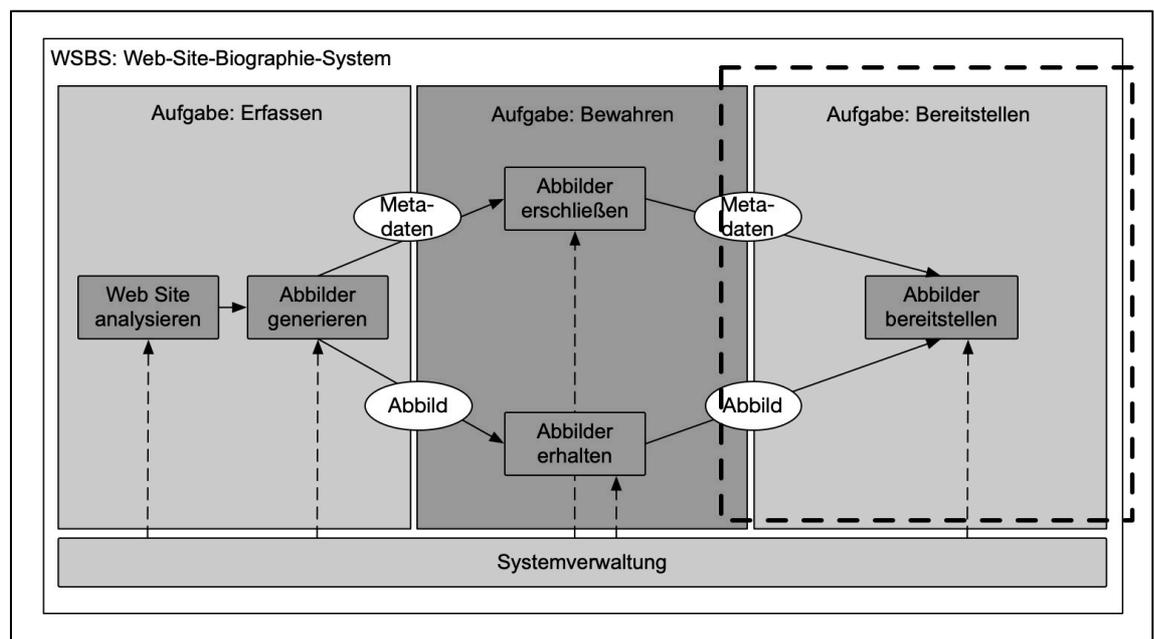


Abb. 6: Bereitstellen im Strukturmodell des WSBS

Um die langfristige Nutzbarkeit der WSB sicherzustellen, muss nicht nur das Modul *Abbilder erhalten* standardisierte Dateiformate verwenden, sondern auch das Modul *Abbilder bereitstellen* diese Dateiformate interpretieren. Abgebildete Inhalte müssen

134 Vgl. Marshall, Brianna H. (Hrsg.): The complete guide to personal digital archiving, Chicago: ALA Editions, an imprint of the American Library Association 2018, S. 41 f.

135 Vgl. Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 59.

durch das WSBS reproduziert und für Anwender nutzbar gemacht werden. Erfolgt die Bewahrung der Abbilder beispielsweise auf Basis des im vorliegenden Kapitel erwähnten WARC-Formats, ist für die Realisierung der Bereitstellungskomponente eine Systemlösung notwendig, die Abbilder im WARC-Format lesen und im Web Browser eines Nutzers reproduzieren kann. Für die Realisierung einer solchen Lösung bestehen zwei Möglichkeiten: Zum einen kann eine Individual-Software entwickelt werden. Zum anderen kann eine bereits erprobte Lösung verwendet werden, die Funktionen zur Reproduktion der Abbilder im WARC-Format bereitstellt. Erprobte Lösungen sind u. a. die Programm-Bibliotheken *pywb* und *OpenWayback*.<sup>136</sup> Diese Programm-Bibliotheken bieten zwar keine umfassenden Funktionen für die Bereitstellung einer WSB, beinhalten aber praxiserprobte Reproduktions-Mechanismen, die als technischer Grundstock für die Entwicklung einer Bereitstellungskomponente verwendet werden können. Durch eine Weiterentwicklung der Funktionsmenge können Bedienoberflächen entstehen, die auch Abbilder eines WSB erlebbar machen.<sup>137</sup> Da sowohl eine Neuentwicklung als auch eine Weiterentwicklung vorab nicht ausgeschlossen werden können, muss diese Entscheidung auf Basis des Systementwurfs erfolgen.

Unabhängig von der technischen Implementierung müssen die Bedienoberflächen zur Bereitstellung einer WSB Abbilder präsentieren und auffindbar machen. Eine Suchfunktion ist in diesem Zusammenhang wesentlich. Bei der Ausgestaltung der Suchfunktion müssen Systemnutzer berücksichtigt werden, die ohne ein konkretes Ziel durch die WSB wandern und Systemnutzer, die ein konkretes Abbild zu einem bestimmten Zeitpunkt suchen.<sup>138</sup> Um das Abbild erlebbar zu machen, ist nicht nur die originalgetreue Reproduktion wesentlich, sondern auch die Darstellung von Kontext-Informationen. Kontext-Informationen sind notwendig, um betriebsspezifische Eigenschaften einer Web Site zu bestimmten Zeitpunkten in der Vergangenheit zu beschreiben. Sie geben einem Betrachter z. B. Aufschluss über die IT-Infrastruktur und über Personal, das in der Vergangenheit an der Erstellung einer Web Site beteiligt war. Eine historische Interpretation der organisationalen Web Site wird mit Hilfe der Kontext-Informationen unterstützt.

---

136 Vgl. Kreymer, Ilya: Pywb Webrecorder web archive replay and capture tools, Online im Internet: <https://pypi.org/project/pywb/>, 13.12.2019. oder Marshall, Brianna H. (Hrsg.): The complete guide to personal digital archiving, a. a. O., S. 41 oder Melo, F.; Gomes, D.; Bicho, D.: A Comparison Between The Performance of Wayback Machines, Online im Internet: <http://sobre.arquivo.pt/sobre/publicacoes-1/a-comparison-between-the-performance-of-wayback>, 13.12.2019.

137 Vgl. zur Aufwandsminimierung Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 154 f.

138 Vgl. Feliciati, Pierluigi; Dobрева, Milena (Hrsg.): Access to digital archives: studying users' expectations and behaviours, in: Digital Archives: Management, use and access, London: Facet Publishing 2018, S. 121 f.

### 3 Eigenschaftsanalyse einer Web Site

#### 3.1 Systematisierung der Eigenschaftsanalyse

Die Beschreibung der Eigenschaften einer Web Site umfasst eine *inhaltliche*, *gestalterische* und *betriebspezifische* Perspektive. Die Aufteilung in den genannten Dreiklang basiert zunächst auf der Komposition einer Web Site in Inhalt, Struktur und Design.<sup>139</sup> Im Rahmen der vorliegenden Arbeit werden die Struktur und das Design aus analytischen Gründen zu einer gestalterischen Perspektive fusioniert.<sup>140</sup>

Als Ergänzung zur *inhaltlichen* und *gestalterischen* Perspektive wird eine *betriebspezifische* Perspektive hinzugefügt. Die betriebspezifische Perspektive berücksichtigt die technische und die organisatorische (IT-)Infrastruktur zum Bertreiben einer Web Site.

Die Kombination aus inhaltlicher, gestalterischer und betriebspezifischer Perspektive ermöglicht eine systematische Strukturierung der Web-Site-Eigenschaften und unterstützt eine stringente Merkmalsbeschreibung des Bezugsobjektes. In der Eigenschaftsanalyse werden fünfzehn Eigenschaften festgelegt, die eine Web Site im Kontext des vorliegenden Konstruktionsziels auszeichnen. Es werden fünf inhaltliche Eigenschaften, vier gestalterische Eigenschaften und sechs betriebspezifische Eigenschaften definiert, die für das Ableiten von Anforderungen unmittelbar relevant sind.

Abbildung 7 zeigt die Eigenschaften einer Web Site einleitend und im Überblick.

Inhaltliche Eigenschaften	<ul style="list-style-type: none"> <li>• Flüchtigkeit</li> <li>• Interaktivität</li> <li>• Absicherung</li> </ul>	<ul style="list-style-type: none"> <li>• Kritikalität</li> <li>• Personalisierung</li> </ul>
Gestalterische Eigenschaften	<ul style="list-style-type: none"> <li>• Grafische Aufbereitung</li> <li>• Positionierung</li> </ul>	<ul style="list-style-type: none"> <li>• Web-Site-Strukturierung</li> <li>• Vielgestaltigkeit</li> </ul>
Betriebspezifische Eigenschaften	<ul style="list-style-type: none"> <li>• Formatvielfalt</li> <li>• Infrastrukturintegration</li> <li>• Infrastrukturwandel</li> </ul>	<ul style="list-style-type: none"> <li>• Verteilung</li> <li>• Sicherheit</li> <li>• Interoperabilität</li> </ul>

Abb. 7: Signifikante Eigenschaften einer Web Site

139 Vgl. Zschau, O.; Traub, D.; Zahradka, R.: Web Content Management. Websites professionell planen und betreiben. 2., überarb. und erw. Aufl, Bonn: Galileo Press 2002, S. 59 f.

140 Vgl. zum Dreiklang auch Krüger, Jörg Dennis; Kopp, Matthias: Web-Content managen: professioneller Einsatz von Content-Management-Systemen, a. a. O., S. 169.

## 3.2 Inhaltliche Eigenschaften

### 3.2.1 Flüchtigkeit

Organisationale Web Sites werden im Regelfall betrieben, um ausgewählte Inhalte zu Produkten, Dienstleistungen, Zielen oder Visionen zu präsentieren.<sup>141</sup> In diesem Zusammenhang müssen Verantwortliche nicht nur entscheiden, zu welchen Themenbereichen Inhalte veröffentlicht werden, sondern auch in welchen Frequenzen das Verändern oder Entfernen dieser Inhalte stattfindet. Die daraus resultierende Flüchtigkeit der Inhalte adressiert die erste inhaltspezifische Eigenschaft.<sup>142</sup>

Die Inhalte einer Web Site sind im Regelfall nicht beständig, sondern weisen unterschiedliche Haltbarkeiten auf. Neben Real-Time-Informationen wie Börsen-Kursen werden in episodischen oder periodischen Abständen z. B. News, Downloads oder Betextungen erstellt, verändert und gelöscht. Während sich einige News-Artikel stündlich oder minutenweise aktualisieren, existieren andere Inhalte über Jahre, ohne Veränderungen zu unterliegen. Hier sind u. a. Inhalte über das Ziel oder die Mission eines Unternehmens anzuführen, ebenso wie Darstellungen der Unternehmensführung oder der Gründungsgeschichte einer Organisation.<sup>143</sup>

Unabhängig davon, wann die Manipulationen einzelner Inhalte stattfinden, sind diese häufig mit einem Verlust ursprünglicher Inhalte verbunden. Zwar verfügen die meisten IT-Systeme zur Verwaltung einer Web Site über Möglichkeiten der Archivierung, diese werden jedoch nicht immer konsequent und umfänglich ausgeschöpft.<sup>144</sup> Es ist nicht auszuschließen, dass Redakteure Inhalte einer Web Site verändern, ohne eine Archivierung vorheriger Zustände durchzuführen. Inhalte werden ausgetauscht und frühere Zustände gehen in Teilen oder gänzlich verloren.<sup>145</sup> Zudem sind Zustände einer Web Site bei einem

---

141 Vgl. Theobald, Elkre; Häusler, Annika: Die Website als zentrales Element der digitalen Markenführung, in: Brand Evolution: moderne Markenführung im digitalen Zeitalter, Hrsg.: Haisch, Philipp T., Wiesbaden: Gabler 2011, S. 220.

142 Vgl. Brügger, Niels: Web History, an Emerging Field of Study, a. a. O., S. 7.

143 Vgl. z. B. Adar, Eytan; Teevan, Jaime; Dumais, Susan T.; Elsas, Jonathan L.: The web changes everything: understanding the dynamics of web content, in: Proceedings of the Second ACM International Conference on Web Search and Data Mining (WSDM 2009): Barcelona, Spain, February 9-12, 2009, S. 284 ff.

144 Vgl. Zschau, O.; Traub, D.; Zahradka, R.: Web Content Management. Websites professionell planen und betreiben. 2., überarb. und erw. Aufl, a. a. O., S. 56.

145 Vgl. Hale, Scott A.; Blank, Grant; Alexander, Victoria D.: Live versus archive: Comparing a web archive to a population of web pages, a. a. O., S. 45. oder Brügger, Niels: Web History, an Emerging Field of Study, a. a. O., S. 7.

Wechsel des zugrundeliegenden IT-Systems gefährdet.<sup>146</sup> Der Migrationsaufwand für das Überführen der Inhalte in neue Systemumgebungen ist häufig zu hoch und ein Transport in neue Systeme wird nicht immer bzw. selten umgesetzt.<sup>147</sup>

### 3.2.2 Interaktivität und Absicherung

Gleichlaufend mit einer zunehmenden Flüchtigkeit der Inhalte hat sich das WWW über Jahre zu einem interaktiven Informationssystem weiterentwickelt. Auf Web Sites werden längst nicht mehr nur statische Inhalte veröffentlicht, sondern eine Vielzahl von interaktiven Elementen angeboten.<sup>148</sup>

Von Suchformularen über dynamische Menüs bis hin zu komplexen Geschäftsprozessen oder virtuellen Welten kann eine Vielzahl an interaktiven Elementen realisiert werden.<sup>149</sup>

Abbildung 7 stellt die Web Site der Suchmaschine Google als Beispiel für die Integration eines interaktiven Suchformulars zum Senden von Suchanfragen dar.

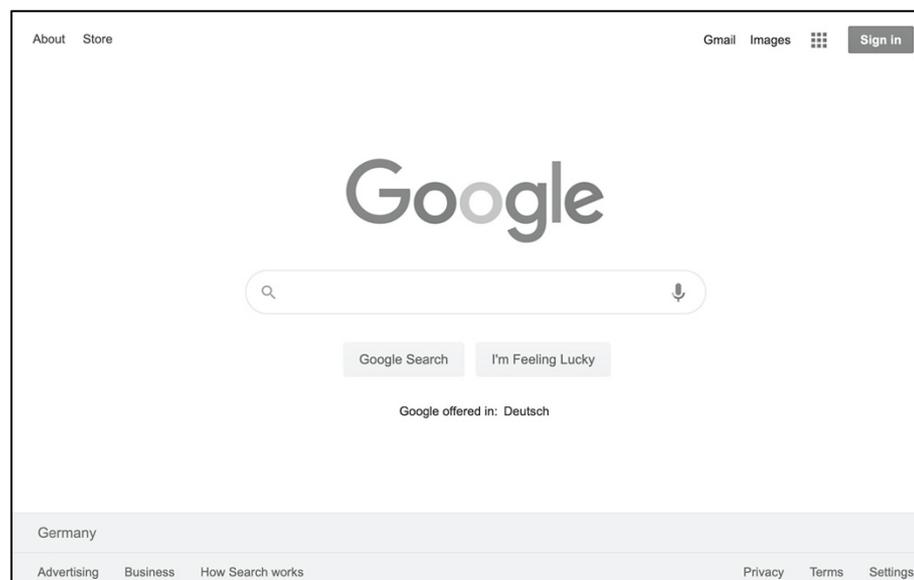


Abb. 8: Web Site der Suchmaschine von Google

146 Vgl. auch Berger, Daniel: Datenverlust: Myspace verliert riesiges Musikarchiv, Online im Internet: <https://www.heise.de/newsticker/meldung/Datenverlust-Myspace-verliert-riesiges-Musikarchiv-4338737.html>, 19.03.2019.

147 Vgl. Masanès, Julien: Web archiving, a. a. O., S. 14 f.

148 Vgl. Chatfield, Tom; Freytag, Carl: 50 Schlüsselideen digitale Kultur, Berlin: Springer Spektrum 2013, S. 68 ff. oder auch Kappel, Gerti; Pröll, Birgit; Reich, Siegfried; Retschitzegger, Werner: Web Engineering – Die Disziplin zur systematischen Entwicklung von Web-Anwendungen, a. a. O., S. 7.

149 Vgl. zur Virtual Reality im Web Krause, Dirk: VR im Web: Virtual-Reality-Web-Apps mit A-Frame programmieren, a. a. O., S. 116 ff.

Ermöglicht wird die Interaktion im Web durch Programme, die Nutzereingaben empfangen, verarbeiten und mit geeigneten Ausgaben reagieren. Typischerweise kommen dabei Programmiersprachen wie JavaScript, PHP, Python oder auch Java zum Einsatz.<sup>150</sup> Besonders die Integration von JavaScript ist allgegenwärtig, auch wenn die Verwendung dieser Skriptsprache aus Sicherheitsbedenken nur begrenzt empfohlen wird.<sup>151</sup>

Abbildung 9 stellt die Integration einer interaktiven Konfigurationsumgebung in die Web Site des Automobilherstellers „Audi“ beispielhaft dar.<sup>152</sup>

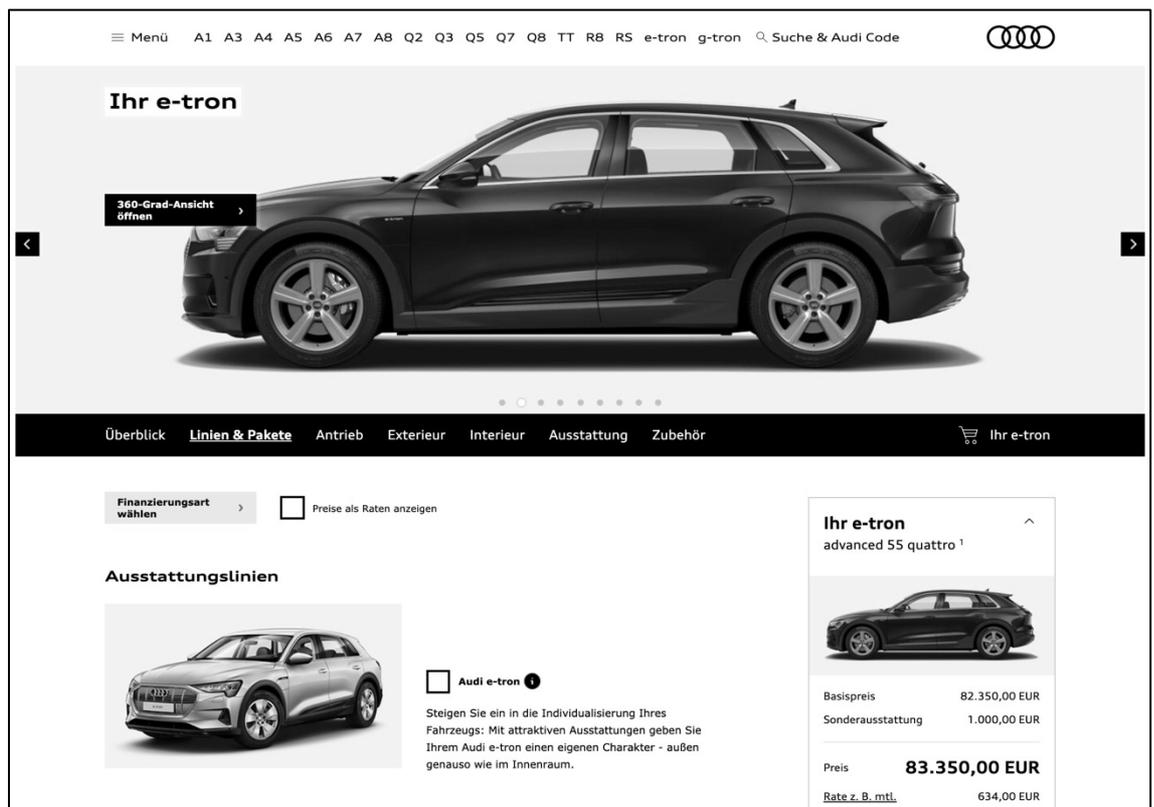


Abb. 9: Online-Konfigurator der Audi AG

Nutzer können ausgewählte Modelle der Marke auf der Unternehmens-Web-Site konfigurieren, visualisieren und gegebenenfalls bestellen.

150 Vgl. Wöhr, Heiko: Web-Technologien: Konzepte - Programmiermodelle - Architekturen, 1. Aufl, Heidelberg: dpunkt-Verl 2004, S. 19 ff. oder Statista (Hrsg.): PYPL. Die beliebtesten Programmiersprachen weltweit laut PYPL-Index im Januar 2018., Online im Internet: <https://de.statista.com/statistik/daten/studie/678732/umfrage/beliebteste-programmiersprachen-weltweit-laut-pypl-index/>, 01.02.2018.

151 Vgl. W3C - World Wide Web Consortium (Hrsg.): The principles of unobtrusive JavaScript, Online im Internet: [https://www.w3.org/wiki/The\\_principles\\_of\\_unobtrusive\\_JavaScript](https://www.w3.org/wiki/The_principles_of_unobtrusive_JavaScript), 03.02.2018.

152 Vgl. AUDI AG (Hrsg.): Audi e-tron, Online im Internet: <https://www.audi.de/de/brand/de/neuwagen/tron/audi-e-tron/linien-pakete.html>, 06.11.2019.

Prinzipiell ist das Nutzbarmachen interaktiver Inhalte in Reproduktionsumgebungen wie Archiven herausfordernd, denn es werden komplexe Mechanismen vorausgesetzt, um die Interaktivität außerhalb der produktiven Web Site erlebbar zu machen.<sup>153</sup> Es muss nicht nur der HTML-Quellcode für einfache Formulare überführt werden, sondern auch Programm-Code für komplexere Inhalte, die auf Basis eigenständiger Software-Lösungen wie Flash oder Java realisiert werden.<sup>154</sup>

Neben der Interaktivität wird eine weitere inhaltsspezifische Eigenschaft durch den Einsatz abgesicherter Inhalte bestimmt. Klassischerweise wird eine Web Site in öffentliche und abgesicherte Bereiche gegliedert. Die Inhalte in abgeschlossenen Bereichen stehen einem Nutzer typischerweise erst nach einer erfolgreichen Anmeldung zur Verfügung.<sup>155</sup>

Abbildung 10 visualisiert die Web-Präsenz des Online-Händlers „Amazon“ und stellt beispielhaft die Anmeldeseite zum Abrufen abgesicherter Inhalte dar.<sup>156</sup>

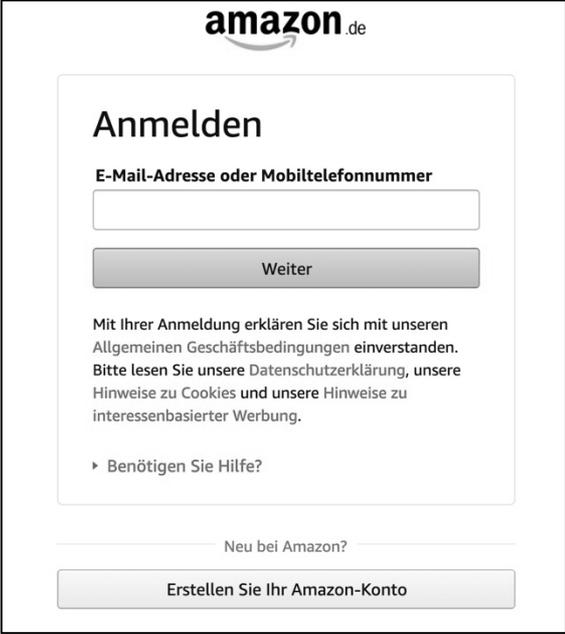


Abb. 10: Anmeldeseite auf amazon.de

153 Vgl. Masanès, Julien: Web Archiving: Issues and Methods, a. a. O., S. 16. oder f. Brown, Adrian: Archiving websites: a practical guide for information management professionals, a. a. O., S. 57.

154 Vgl. auch Noak, Pit: Die Zukunft der Archive: Games nicht nur bewahren, sondern spielbar erhalten, a. a. O., 29.12.2018.

155 Vgl. Schwickert, Axel C.: Web Site Engineering: Ökonomische Analyse und Entwicklungssystematik für eBusiness-Präsenzen, a. a. O., S. 93 ff.

156 Vgl. Amazon.com, Inc (Hrsg.): Amazon.de: Günstige Preise für Elektronik & Foto, Filme, Musik, Bücher, Games, Spielzeug & mehr, Online im Internet: <https://amazon.de>, 03.01.2020.

### 3.2.3 Kritikalität und Personalisierung

Sowohl in öffentlichen als auch in abgeschlossenen Segmenten einer Web Site werden Inhalte veröffentlicht, die gesetzlichen Regelungen unterliegen können. Typische Gesetze finden sich im Datenschutz, Urheberrecht oder Verwertungsrecht.<sup>157</sup>

Ein prominentes Beispiel für den Einfluss gesetzlicher Regelungen auf Inhalte ist die Veröffentlichung von Fotos, die Mitarbeiter eines Unternehmens präsentieren. Abbildungen einer Person sind nur nach Einwilligung der betroffenen Person zulässig. Nach dem Ausscheiden bzw. nach der Beendigung des Arbeitsverhältnisses sind Fotos des Mitarbeiters von der Web Site zu entfernen. Geschieht dies nicht, greifen Regelungen, die zum Schutz des Persönlichkeitsrechts verfasst wurden.<sup>158</sup>

Neben den Persönlichkeitsrechten ist auch das Urheberrecht von unmittelbarer Relevanz. Die Verwendung von Texten, Bildern, Videos, Software und anderen Inhalten ist im Rahmen eines geeigneten Risiko-Managements zu bewerten. Nicht nur zum Zeitpunkt der Veröffentlichung muss das Urheberrecht gewahrt werden, sondern auch bei der langfristigen Dokumentation in externen oder internen Sicherungsumgebungen.<sup>159</sup> Im Idealfall geht die Aufbewahrung kritischer Inhalte mit der Einwilligung der Inhaltseigentümer einher. Ist dies nicht der Fall, müssen Inhalte, die beispielsweise das Urheberrecht verletzen, jederzeit eliminiert werden können, um Schadensersatzklagen entgegenzuwirken.<sup>160</sup>

Ähnliche Konsequenzen gilt es im Zusammenhang mit Regelungen des Datenschutzes zu vermeiden.<sup>161</sup> Längst hat die im Jahr 2016 beschlossene „Europäische Datenschutz-Grundverordnung (EU-DSGVO)“ Einfluss auf Unternehmen im europäischen Rechtsraum genommen. Auch Web-Site-Betreiber sind betroffen und stehen vor einer nicht unerheblichen Herausforderung. Organisationen sind gemäß EU-DSGVO verpflichtet, personenbezogene Daten nur mit expliziter Genehmigung zu erfassen und zu speichern. Es

---

157 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 51.

158 Vgl. in diesem Zusammenhang dejure.org Rechtsinformationssysteme GmbH (Hrsg.): Rechtsprechung LAG Hessen, 24.01.2012 - 19 SaGa 1480/11, Online im Internet: <https://dejure.org/2012,733,03.02.2018>.

159 Vgl. z. B. Brown, Adrian: Archiving websites: a practical guide for information management professionals, a. a. O., S. 147.

160 Vgl. Brown, Adrian: Archiving websites: a practical guide for information management professionals, a. a. O., S. 152.

161 Vgl. Brown, Adrian: Archiving websites: a practical guide for information management professionals, a. a. O., S. 152 f.

gilt grundsätzlich ein Verbot zur Verarbeitung personenbezogener Daten mit Erlaubnisvorbehalt.<sup>162</sup> Darüber hinaus wurde auch die Pflicht zum Löschen digitaler Daten angepasst. Die EU-DSGVO sieht z. B. vor, dass personenbezogene Daten unverzüglich eliminiert werden, sobald der Zweck ihrer Verarbeitung nicht mehr gegeben ist. Grundsätzlich sind entsprechende Anpassungen an den Datenschutz für die Regelkonformität eines Unternehmens unerlässlich. Die organisatorische und technische Implementierung geeigneter Maßnahmen kann jedoch mit erheblichem Mehraufwand verbunden sein, stellt im Zweifel aber die Geschäftsfähigkeit sicher.<sup>163</sup>

Setzt man das Konstruktionsziel der vorliegenden Arbeit in Relation, ist festzustellen, dass nicht nur die Web Site an sich, sondern auch deren Abbilder in einer Speicherumgebung von entsprechenden Datenschutzregelungen betroffen sind. Sobald kritische Inhalte von einer Web Site eliminiert werden müssen, sind diese auch aus historischen Präsentationsumgebungen zu entfernen.<sup>164</sup>

Eine Menge dieser kritischen Inhalte ist zudem personalisiert. Programme, die auf den Servern der Web-Site-Betreiber installiert sind, analysieren nicht selten das Verhalten einzelner Nutzer und generieren Nutzer-bezogene Inhalte. Beispielhaft können an dieser Stelle Online-Shops mit individuellen Produktvorschlägen angeführt werden. Auf Basis des Klick- und Kaufverhaltens eines Kunden werden personalisierte Angebote vorgeschlagen, die ein angepasstes Einkaufserlebnis ermöglichen.<sup>165</sup>

Empfehlungssysteme erfassen die Eingaben der Nutzer, verarbeiten diese mit geeigneten Algorithmen und generieren einen Output, der sich aus dem Gelernten speist. Durch diese Systeme kann für Nutzer eine angepasste Benutzeroberfläche im Web generiert werden, deren grundlegende Struktur und Darstellung genauso variieren kann wie die darauf präsentierten Inhalte. Anwendungsdomänen der Empfehlungssysteme sind vielfältig und Möglichkeiten der Personalisierung mit erheblichen Potentialen im Rahmen des Produkt-Marketings und der Optimierung von Kaufprozessen verbunden.<sup>166</sup>

---

162 Vgl. Art. 6 EU-DSGVO im Amtsblatt der Europäischen Union (Hrsg.): VERORDNUNG (EU) 2016/679 DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG, Online im Internet: <http://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32016R0679&from=DE>, 05.02.2018. Als Ergänzung vgl. Akinici, Nick; Heidrich, Joerg: Informationsverpflichtet, in: iX, 1/2018, S. 38 f.

163 Vgl. hierzu Heidrich, Jörg: Pflichtverschiebung, in: iX, 1/2018, S. 46.

164 Vgl. Brown, Adrian: Archiving websites: a practical guide for information management professionals, a. a. O., S. 152 ff.

165 Vgl. Hassler, Marco: Web Analytics: Metriken auswerten, Besucherverhalten verstehen, Website optimieren, a. a. O., S. 46 f.

166 Vgl. Klahold, André: Empfehlungssysteme: Recommender Systems – Grundlagen, Konzepte und Lösungen, 1. Aufl, Wiesbaden: Vieweg + Teubner 2009, S. 4 ff.

Eine vergleichsweise einfache Form der Personalisierung findet zudem durch das manuelle Bereitstellen von Personen-bezogenen Inhalten statt. Neben der automatisierten Personalisierung auf Basis von Empfehlungssystemen werden häufig manuell erfasste Inhalte wie Bank-Daten oder Kontoinformationen als Bestandteil einer Web Site gespeichert und gepflegt.

Abbildung 11 stellt Menüpunkte zum manuellen Verwalten von personalisierten Inhalten dar, die ein Nutzer im Online Shop von Amazon erstellen und pflegen kann.<sup>167</sup>

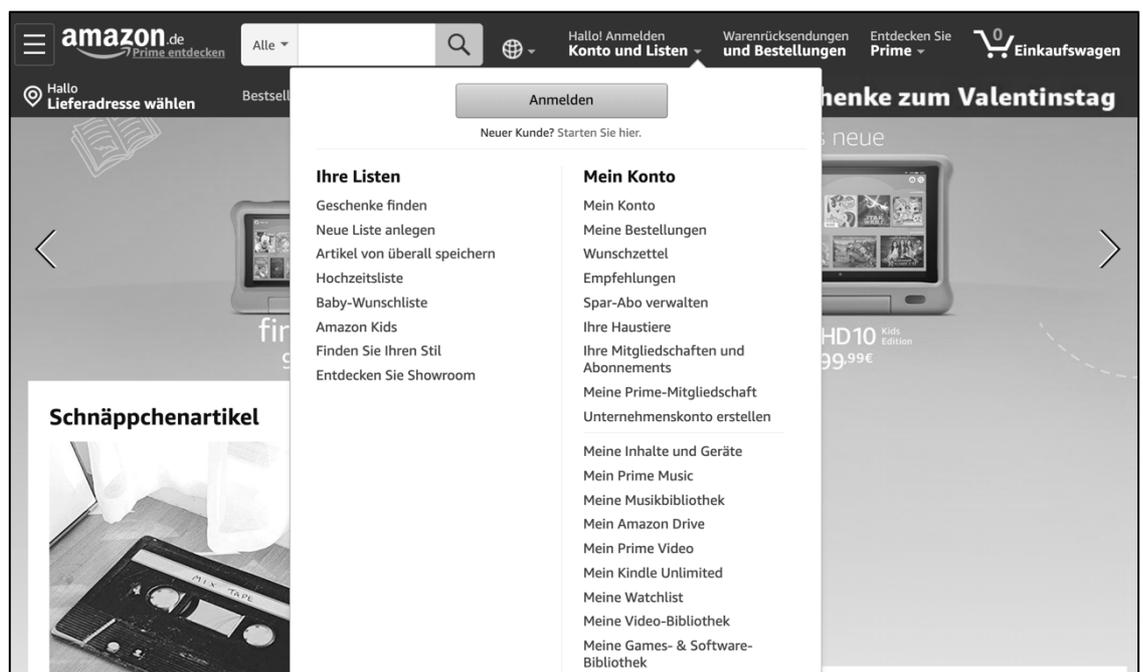


Abb. 11: Menü zum Verwalten von personalisierten Inhalten auf amazon.de

Nachdem im vorliegenden Kapitel inhaltliche Merkmale einer Web Site aufgeführt wurden, folgt im nachstehenden Kapitel die Definition und Erläuterung gestalterischer Eigenschaften einer Web Site.

<sup>167</sup> Vgl. Amazon.com, Inc (Hrsg.): Amazon.de: Günstige Preise für Elektronik & Foto, Filme, Musik, Bücher, Games, Spielzeug & mehr, a. a. O., 03.01.2020.

### 3.3 Gestalterische Eigenschaften

#### 3.3.1 Grafische Aufbereitung und Positionierung

Im Regelfall entscheiden sich die Betreiber einer Web Site dazu, Inhalte systematisch und mit Hilfe geeigneter Informationsarchitekturen zu verteilen.<sup>168</sup> In diesem Zusammenhang werden Positionierungen der Inhalte festgelegt und grafische Aufbereitungen vorgenommen. Die erste gestalterische Eigenschaft adressiert die grafische Aufbereitung einer Web Site.

Auf einer Web Site lassen sich Farben für Texte, Links oder Umrandungen definieren, Hintergründe gestalten und individuelle Schriftarten bestimmen. Aufbauend auf einer einfachen HTML-Datei mit strukturierten Text-Elementen kann eine ansprechende Oberfläche gestaltet werden.<sup>169</sup> Durch eine geeignete Farbgestaltung werden Texte oder Funktionen hervorgehoben, um u. a. einer Informationsüberlastung der Nutzer vorzubeugen.<sup>170</sup> Darüber hinaus kann die Corporate Identity problemlos in den digitalen Raum transportiert werden, denn die Web Site lässt sich durch grafische Befehle an das allgemeine Erscheinungsbild der Organisation anpassen. Farben und Formen orientieren sich an etablierten grafischen Elementen, wie dem Firmenlogo.<sup>171</sup>

Zur grafischen Aufbereitung werden sogenannte „Style Sheets“ verwendet. Style Sheets sind Dateien, in denen (im Idealfall alle) grafische Eigenschaften einer Web Site definiert werden.<sup>172</sup> In Ergänzung zu diesen Dateien ist das Eingliedern von Schriftarten gegenwärtig, die von standardisierten Schriftarten des Web Browsers abweichen. Neben klassischen Schriftarten wie „Arial“ existiert eine Vielzahl von Web-Schriftarten, die durch das Einbinden von zusätzlichen Dateien genutzt werden können. Einer der wohl bekanntesten Anbieter für Web-Schriftarten ist Google. Unter <https://fonts.google.com> werden diverse Schriftarten zum Laden und Verwenden auf Web Sites angeboten.<sup>173</sup>

---

168 Vgl. Bühler, Peter; Schlaich, Patrick; Sinner, Dominik: Webdesign: Interfacedesign - Screendesign - Mobiles Webdesign, Berlin: Springer Vieweg 2017, S. 21 ff.

169 Vgl. W3C - World Wide Web Consortium (Hrsg.): WHAT IS CSS?, Online im Internet: <https://www.w3.org/Style/CSS/>, 10.02.2018.

170 Vgl. Meidl, Oliver: Global Website: Webdesign im internationalen Umfeld, Wiesbaden: Springer Gabler 2013, S. 22.

171 Vgl. Meidl, Oliver: Global Website: Webdesign im internationalen Umfeld, a. a. O., S. 19.

172 Vgl. Bühler, Peter; Patrick Schlaich; Sinner, Dominik: HTML5 und CSS3: Semantik - Design - Responsive Layouts, Berlin, Germany: Springer Vieweg 2017, S. 45.

173 Vgl. Bühler, Peter; Patrick Schlaich; Sinner, Dominik: HTML5 und CSS3: Semantik - Design - Responsive Layouts, a. a. O., S. 60.

Ein Großteil der Web Sites wird jedoch nicht nur durch Style Sheets und Schriftarten grafisch aufbereitet, sondern an Endgeräte, Zielgruppen oder Gestaltungs-Trends angepasst. Inhalte werden häufig nach Bildschirmgröße und Seitenverhältnis positioniert.<sup>174</sup> Dabei ist ein Dreiklang festzustellen, der sich an Auflösungen von Arbeitsplatzrechnern, Tablets und Smartphones orientiert.<sup>175</sup>

Abbildung 12 stellt die Gestaltungsversionen der Web Site des Automobilherstellers „BMW“ für Arbeitsplatzrechner und Smartphones dar.

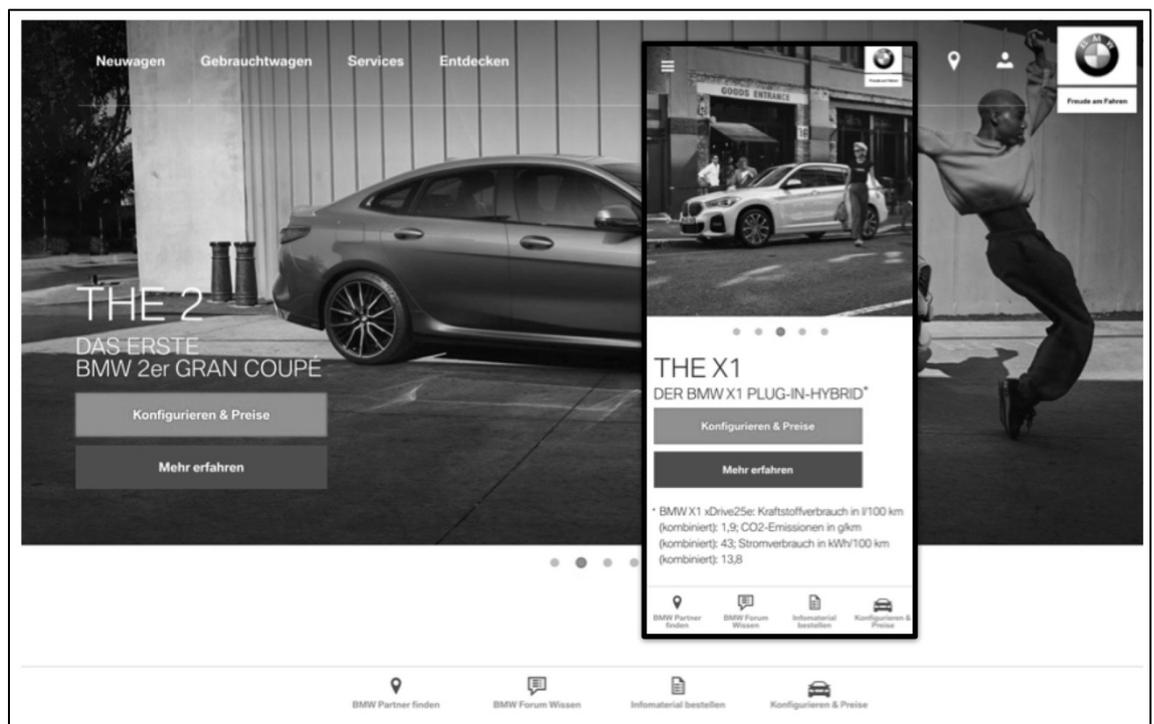


Abb. 12: Gestaltungsansätze der BMW-Web-Site nach Endgeräten

Für die automatische Anpassung einer Web Site an Endgeräte sind verschiedene technische Konzepte beobachtbar. Während einige Web Sites durch prozentuale Aufteilungen der Inhalte charakterisiert sind, verändern andere Web Sites die Anzahl und Menge der darzustellenden Elemente. Je nach Bildschirmgröße wird eine passende Version der

174 Vgl. auch hier Bühler, Peter; Schlaich, Patrick; Sinner, Dominik: Webdesign: Interfacedesign - Screendesign - Mobiles Webdesign, a. a. O., S. 47.

175 Vgl. Bühler, Peter; Schlaich, Patrick; Sinner, Dominik: Webdesign: Interfacedesign - Screendesign - Mobiles Webdesign, a. a. O., S. 47 ff. oder Thesmann, Stephan: Interface design: usability, user experience und accessibility im Web gestalten, 2., aktualisierte und erweiterte Auflage, Wiesbaden: Springer Vieweg 2016, S. 259 f.

Web Site geladen. Hybride Gestaltungsansätze bedienen sich beider Konzepte.<sup>176</sup> Bei einer Veränderung werden Anpassungen zunächst durch prozentuale Aufteilungen realisiert, sobald bestimmte Auflösungen unter- oder überschritten werden, können Inhalt eliminiert oder nachgeladen werden.<sup>177</sup>

### 3.3.2 Web-Site-Strukturierung und Vielgestaltigkeit

Ergänzend zur grafischen Aufbereitung und Positionierung sind differierende Strukturen zur Navigation charakteristisch für eine Web Site. Grundlage für das Erzeugen der Navigationsstrukturen sind Hyperlinks (kurz: Links). Links werden als Bestandteile von Menüs oder z. B. im Freitext eingepflegt, um einzelne Web-Seiten zu verbinden.<sup>178</sup>

Abbildung 13 stellt typische Navigationsstrukturen einer Web Site schematisch dar.<sup>179</sup>

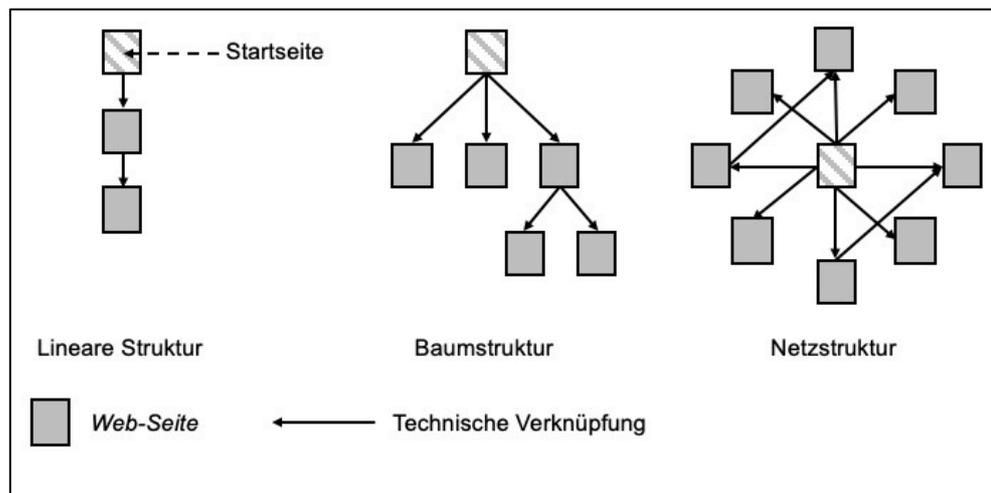


Abb. 13: Visualisierung der Strukturen einer Web Site

Klassische Medien wie Bücher oder Filme sind durch lineare Navigationsstrukturen gekennzeichnet. Nutzer beginnen typischerweise mit der ersten Seite eines Buches und durchblättern dieses sequentiell, Seite für Seite. Ein Film startet üblicherweise bei der ersten Filmminute und ist durch eine definierte Abfolge von Sequenzen zur Inhaltsvermittlung charakterisiert. Diese linearen Strukturen der Inhaltsvermittlung sind längst

176 Vgl. Carlos, Gilberto: Responsive Web Design with jQuery, Birmingham: Packt Publishing 2013, S. 12 ff.

177 Vgl. Bühler, Peter; Schlaich, Patrick; Sinner, Dominik: Webdesign: Interfacedesign - Screendesign - Mobiles Webdesign, a. a. O., S. 49 f.

178 Vgl. Bühler, Peter; Schlaich, Patrick; Sinner, Dominik: Webdesign: Interfacedesign - Screendesign - Mobiles Webdesign, a. a. O., S. 26 ff.

179 Vgl. hierzu Bühler, Peter; Schlaich, Patrick; Sinner, Dominik: Webdesign: Interfacedesign - Screendesign - Mobiles Webdesign, a. a. O., S. 21 ff.

etabliert, lassen sich aber nicht immer auf Web Sites übertragen.<sup>180</sup> Im Regelfall existieren unterschiedliche Pfade, die unterschiedliche Zielgruppen adressieren. Neben linearen Navigationsstrukturen entstehen auf einer Web Site Baum- oder Netzstrukturen.<sup>181</sup>

Vergleichbar mit der Flüchtigkeit von Inhalten unterliegt sowohl die Navigationsstruktur als auch die grafische Aufbereitung und Positionierung einer Web Site einem ständigen Wandel. Neben der Anpassung an Endgeräte ist eine saisonale oder ereignisgebundene Umgestaltung der Web-Präsenz einer Organisation beobachtbar.<sup>182</sup> Organisationen verändern die Gestalt ihrer Web Site, um in der Menge an virtuellen Angeboten nicht zu verschwinden und für die Kunden attraktiv zu bleiben.<sup>183</sup>

Seit den Ursprüngen des Internets hat sich das WWW rasant und beständig weiterentwickelt. Die Anfänge des Internet-Dienstes waren durch unformatierte Texte charakterisiert. Mit dem Aufkommen der ersten Web Browser begann jedoch eine kontinuierliche Weiterentwicklung der Gestaltungsansätze.<sup>184</sup> Neu aufkeimende Web-Technologien schaffen immer komplexere Gestaltungsmöglichkeiten, die Web Sites längst für mobile Endgeräte oder auch als Spieleplattform nutzbar machen.<sup>185</sup>

Gestaltungselemente verändern sich aber nicht nur über die Zeit, sondern variieren auch je nach Segment einer Web Site. Das öffentliche Segment einer Web Site ist z. B. häufig durch das Darstellen von Produkten oder Dienstleistungen charakterisiert und dient als Instrument der Markenführung.<sup>186</sup> Im Rahmen der Online-Kommunikation stehen unter anderem Produktpräsentationen, lesefreundliche Texte und eine geeignete Informationsarchitektur im Fokus.<sup>187</sup>

Neben der öffentlichen Web-Präsenz existiert oftmals auch ein abgeschlossener Bereich. Im Regelfall werden im abgeschlossenen Bereich der Web Site Funktionen angeboten,

---

180 Vgl. Bühler, Peter; Schlaich, Patrick; Sinner, Dominik: Webdesign: Interfacedesign - Screendesign - Mobiles Webdesign, a. a. O., S. 21 f.

181 Vgl. Hassler, Marco: Web Analytics: Metriken auswerten, Besucherverhalten verstehen, Website optimieren, a. a. O., S. 411.

182 Vgl. Sens, Bastian: Schluss mit 08/15-Websites – so bringen Sie Ihr Online-Marketing auf Erfolgskurs: 79 Tipps für Ihren Online-Auftritt, Wiesbaden: Springer Gabler 2017, S. 14 ff.

183 Vgl. Schüle, Hubert: Die Usability von Websites, a. a. O., S. 1092.

184 Vgl. Hansch, Pierre; Rentschler, Christian: EmotionWeb: emotionale Websites durch Bewegtbild und Sound-Design, Berlin: Springer Vieweg 2012, S. 15 f. und Weitekamp, Lea: JavaScript, Flash, CSS und mehr: Die Geschichte des Webdesigns in 9 GIFs, Online im Internet: <https://t3n.de/news/javascript-flat-geschichte-webdesign-585903/>, 16.02.2018.

185 Weitekamp, Lea: JavaScript, Flash, CSS und mehr: Die Geschichte des Webdesigns in 9 GIFs, a. a. O., 16.02.2018.

186 Vgl. Theobald, Elkre; Häusler, Annika: Die Website als zentrales Element der digitalen Markenführung, a. a. O., S. 219 ff.

187 Vgl. Kilian, Thomas; Langner, Sascha: Online-Kommunikation: Kunden zielsicher verführen und beeinflussen, 1. Aufl, Wiesbaden: Gabler 2010, S. 23.

die das Durchführen von Bestellprozessen oder eine Verwaltung und Verwendung von Nutzerprofilen ermöglichen. Als Beispiele zur Verwendung von Nutzerprofilen können Chats oder Foren zur Kommunikation genannt werden. Zur benutzerfreundlichen Darstellung der Funktionen werden im Regelfall Positionierungen bzw. grafische Aufbereitungen forciert, die von den Gestaltungsansätzen der öffentlichen Web Site abweichen.<sup>188</sup>

Aufbauend auf den vorgenannten gestalterischen Eigenschaften einer Web Site, folgt im nachstehenden Kapitel die Definition betriebspezifischer Eigenschaften.

### 3.4 Betriebsspezifische Eigenschaften

#### 3.4.1 Formatvielfalt

Inhalt und Gestalt einer Web Site sind das Ergebnis technischer und organisatorischer Operationen, die auf Basis einer geeigneten Infrastruktur realisiert werden.<sup>189</sup> Damit die Lebensgeschichte einer Web Site erlebbar gemacht werden kann, müssen diese betriebspezifischen Eigenheiten analysiert und dokumentiert werden. Nur unter Berücksichtigung der Betriebsumgebung können umfassende Interpretationen der organisationalen Web-Site-Vergangenheit abgeleitet werden.<sup>190</sup>

Als erste betriebspezifische Eigenschaft wird die Vielfalt der Dateiformate (Formatvielfalt) betrachtet. Neben dem klassischen Hypertext-Dateiformat, welches für das Erzeugen einer Web Site grundlegend ist, wird eine Vielzahl an weiteren Dateiformaten zum Erstellen einer Web Site verwendet.<sup>191</sup> Die Dateiformate unterscheiden sich dabei typischerweise nicht nur in der Struktur ihrer Datenströme, sondern setzen unterschiedlichste Programme zum Erstellen und Interpretieren voraus.<sup>192</sup>

Klassische Vertreter zum Darstellen von Bildern sind z. B. die Dateiformate „JPEG“, „GIF“, „TIFF“ oder „PNG“.<sup>193</sup> Auch bei der Einbindung multimedialer Elemente ist eine Vielzahl von Dateiformaten präsent. Es finden seit langer Zeit Dateiformate wie „MP4“,

---

188 Vgl. auch hier Galitz, Wilbert O.: *The essential guide to user interface design: an introduction to GUI design principles and techniques*, 3rd ed, Indianapolis, IN: Wiley Pub 2007, S. 258.

189 Vgl. Zschau, O.; Traub, D.; Zahradka, R.: *Web Content Management. Websites professionell planen und betreiben*. 2., überarb. und erw. Aufl, a. a. O., S. 59 f.

190 Vgl. *Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften, DARIAH-DE* (Hrsg.): *Handbuch Digital Humanities: Anwendungen, Forschungsdaten und Projekte / DARIAH-DE*, Berlin: epubli 2015, S. 66.

191 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): *Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644*, a. a. O., S. 12 f.

192 Vgl. z. B. Ziehl, Stefan: *Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke*, a. a. O., S. 16 f.

193 Vgl. Bühler, Peter; Patrick Schlaich; Sinner, Dominik: *HTML5 und CSS3: Semantik - Design - Responsive Layouts*, a. a. O., S. 22.

„OGG“ oder „WebM“ Verwendung. Das Abspielen audio-visueller Elemente erfolgt durch Flash-Komponenten, Microsoft Silverlight oder HTML-5-Implementierungen.<sup>194</sup> Vergleichbares kann für das Einbinden von Audio-Elementen festgestellt werden. Auch hier existiert eine Menge an Dateiformaten, aus der zur Implementierung einer Audiodatei gewählt werden kann.<sup>195</sup>

Grundsätzlich stellt die Vielfalt an verfügbaren und verwendeten Dateiformaten digitale Erhaltungsinitiativen vor Herausforderungen.<sup>196</sup> Besonders bei Dateiformaten, die nur mit zusätzlicher Software verwendet werden können, sind aufwendige Reproduktionsumgebungen notwendig. Darüber hinaus werden Dateiformate kontinuierlich erweitert und ältere Formate sind nur auf Umwegen für die Nachwelt interpretierbar.<sup>197</sup>

### 3.4.2 Infrastrukturintegration und Infrastrukturwandel

Typischerweise sind nicht nur Dateiformate einer Web Site durch eine begrenzte Haltbarkeit geprägt, sondern auch die zugrundeliegende Infrastruktur. Die Web Site einer Organisation wird im Regelfall nicht alleinstehend betrieben, sondern in eine vorhandene Infrastruktur integriert. Die Integration findet dabei sowohl auf organisatorischer als auch auf technischer Ebene statt.

Mitarbeiter werden mit der Pflege und Erstellung von Inhalten betraut und Betriebs-Komponenten in die bestehende technische Infrastruktur eingegliedert. Die technische Infrastruktur umfasst dabei üblicherweise die Server zur Bereitstellung einer Web Site, Speichersysteme, Netzwerktechnik, Geräte zum Betrieb der Hardware sowie immaterielle Komponenten. Zu den immateriellen Komponenten zählt systemnahe Software wie ein Betriebssystem, Software zur Virtualisierung der Infrastruktur-Komponenten oder Anwendungssysteme (AWS) zum Erstellen der Web Site.<sup>198</sup>

---

194 Vgl. Hansch, Pierre; Rentschler, Christian: *EmotionWeb: emotionale Websites durch Bewegtbild und Sound-Design*, a. a. O., S. 21 ff.

195 Vgl. zu Audioformaten Hansch, Pierre; Rentschler, Christian: *EmotionWeb: emotionale Websites durch Bewegtbild und Sound-Design*, a. a. O., S. 29 ff.

196 Vgl. auch Noak, Pit: *Die Zukunft der Archive: Games nicht nur bewahren, sondern spielbar erhalten*, a. a. O., 29.12.2018.

197 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): *Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644*, a. a. O., S. 12 f. oder zu applikationsgebundenen Web Content Zschau, O.; Traub, D.; Zahradka, R.: *Web Content Management. Websites professionell planen und betreiben*. 2., überarb. und erw. Aufl, a. a. O., S. 41.

198 Vgl. Office of Government Commerce (OGC) (Hrsg.): *The official introduction to the ITIL service lifecycle*, London: Stationary Office 2007, S. 199.

Im Regelfall werden für Web Sites solche AWS verwendet, die eine möglichst nutzerfreundliche Entwicklung und Pflege ermöglichen. Da die Programmkomponenten der AWS meistens einer kontinuierlichen Aktualisierung durch den Hersteller unterliegen, sind Organisationen darauf angewiesen, mit regelmäßigen Update-Zyklen die Betriebsfähigkeit sicherzustellen.<sup>199</sup> Neben einfachen Aktualisierungen sind aber auch Szenarien gegenwärtig, in denen die Hersteller der AWS gewechselt und damit neue Versionen der Web Site implementiert werden. Mit dem Systemwechsel verbundene Veränderungen sind dabei durchaus als Meilenstein in der Lebensgeschichte einer Web Site zu verstehen.

Gerade im Kontext der Erhaltung wird bei einer technischen Infrastruktur besonders auf den Verfall der Speichermedien eingegangen.<sup>200</sup> Mit der begrenzten Haltbarkeit von digitalen Datenträgern ist das Risiko eines Datenverlusts akut und Aktualisierungen der Speicherumgebungen erforderlich. Neben Veränderungen der Speicherumgebungen sind Veränderungen der Betriebssysteme und Rechenleistungen die Regel.<sup>201</sup>

In den Anfängen des WWW wurden beispielsweise Infrastruktur-Komponenten verwendet, deren technische Attribute heute längst nicht mehr ausreichen, um eine organisationale Web Site performant zu betreiben. Getrieben durch die wachsenden Anforderungen an den Funktionsumfang im Web, verbesserte Netzanbindungen und sinkende Hardware-Preise sind regelmäßige Aktualisierungen technischer Komponenten allgegenwärtig.<sup>202</sup>

Für eine stringente Beschreibung einer Web Site müssen neben der technischen Infrastruktur aber auch institutionelle Eigenschaften berücksichtigt werden.<sup>203</sup> Klassischerweise wird die Web Site einer Organisation von einer Menge an Mitarbeitern erstellt, gepflegt und gegebenenfalls auch funktional erweitert.<sup>204</sup> Es ist nicht untypisch, dass Aufgaben, Verantwortlichkeiten oder Kompetenzen eines Unternehmens durch die Web Site beeinflusst werden und über die Jahre wechseln.<sup>205</sup> An dieser Stelle kann wechselndes

---

199 Vgl. z. B. auch hier Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften, DARIAH-DE (Hrsg.): Handbuch Digital Humanities: Anwendungen, Forschungsdaten und Projekte / DARIAH-DE, a. a. O., S. 61 f.

200 Vgl. Schneider, Holger: Digitale Amnesie: Langzeitarchivierung digitaler Dokumente im betrieblichen Umfeld, a. a. O., S. 101.

201 Vgl. Schwenck, Rebecca: Langzeitarchivierung digitaler Informationen: Ein internationaler Überblick über Organisation und Projektarbeit, a. a. O., S. 15.

202 Vgl. Chatfield, Tom; Freytag, Carl: 50 Schlüsselideen digitale Kultur, a. a. O., S. 26.

203 Vgl. Rudolph, Simone: Servicebasierte Planung und Steuerung der IT-Infrastruktur im Mittelstand: ein Modellansatz zur Struktur der IT-Leistungserbringung, Wiesbaden: Gabler 2009, S. 14 ff.

204 Vgl. Heinrich, Lutz J.; Heinzl, Armin; Roithmayr, Friedrich: Wirtschaftsinformatik-Lexikon: mit etwa 4000 Stichwörtern und 3700 Verweisstichwörtern, 7., vollständig überarbeitete und erw. Aufl., München: Oldenbourg 2004, S. 15 f.

205 Vgl. z. B. Schwickert, Axel C.: Web Site Engineering: Ökonomische Analyse und Entwicklungssystematik für eBusiness-Präsenzen, a. a. O., S. 129 ff.

Personal zur Administration der Server oder zum Pflegen der Inhalte genannt werden. Diese organisatorischen Aspekte gilt es auch als Bestandteil der WSB zu dokumentieren.

### 3.4.3 Verteilung, Sicherheit und Interoperabilität

Nicht nur globale Web Sites stehen vor der Herausforderung möglichst geringe Ladezeiten, wenige Ausfälle und andauernde Verfügbarkeit zu gewährleisten, sondern auch regionale bzw. kleinere Organisationen sind bestrebt, mit ihrer Web Site entsprechende Qualitätsmerkmale zu erfüllen.<sup>206</sup> Um eine bestmögliche Verfügbarkeit zu erreichen, werden einzelne Komponenten auf unterschiedliche Hard- oder Software verteilt.<sup>207</sup> Durch diese Verteilung kann z. B. zu Stoßzeiten eine Entlastung der organisationalen Infrastruktur erreicht werden. Die Verteilung einzelner Systembausteine erhöht die Performance der Betriebskomponenten und verringert die Ladezeiten.<sup>208</sup>

In Ergänzung zur Verteilung von Nutzeranfragen werden einzelne Skripte, Grafiken oder Inhalte von externen Dienstleistern eingebunden und zur Gestaltung verwendet. Typische Beispiele sind Bilder oder Programmbibliotheken, die von einem Drittanbieter entwickelt und im Quellcode der Web Site referenziert werden.<sup>209</sup> Darüber hinaus wird die Performance einer Web Site im Regelfall durch das Anwenden geeigneter Sicherheitsmechanismen unterstützt bzw. dauerhaft sichergestellt.

Das Speichern von sensiblen Nutzerdaten oder Unternehmensinformationen in den Datenbanken einer Web Site ist gegenwärtig und ein Schutz dieser Informationen fortwährend akut. Vor allem um die Geschäftsfähigkeit einer Organisation sicherzustellen und damit das Bestehen im digitalen Wirtschaftsgefüge zu gewährleisten.<sup>210</sup>

---

206 Vgl. Meidl, Oliver: Global Website: Webdesign im internationalen Umfeld, a. a. O., S. 52.

207 Vgl. Hassler, Marco: Web Analytics: Metriken auswerten, Besucherverhalten verstehen, Website optimieren, a. a. O., S. 59.

208 Vgl. Bengel, Günther; Baun, Christian; Kunze, Marcel; Stucky, Karl-Uwe (Hrsg.): Masterkurs parallele und verteilte Systeme: Grundlagen und Programmierung von Multicore-Prozessoren, Multiprozessoren, Cluster, Grid und Cloud, 2. erw. und aktualisierte Auflage, Wiesbaden: Springer Vieweg 2015, S. 442.

209 Vgl. z. B. jQuery Foundation (Hrsg.): jQuery – write less, do more, Online im Internet: <https://jquery.com/>, 11.05.2018.

210 Vgl. Bhasin, Shweta: Web security basics, Cincinnati, Ohio: Premier Press 2003, S. 4 f.

Für die Sicherheit einer Web Site müssen unterschiedliche Mechanismen implementiert werden, die auf eine Vielzahl von Angriffsszenarien reagieren.<sup>211</sup> Beispielhaft sind an dieser Stelle Angriffsszenarien wie „SQL-Injections“, „Cross-Site Scripting Attacks (XSS-Attacks)“, „Denial of Service Attacks (DoS-Attacks)“ oder „Distributed Denial of Service Attacks (DDoS-Attacks)“ zu nennen. Die hier genannten Angriffsszenarien verdeutlichen exemplarisch ein Set an Risikofaktoren, das auf eine Web Site wirkt.<sup>212</sup> Darüber hinaus wird heute eine Verschlüsselung von Informationen, die über das Web transportiert werden, gefordert. Informationen bzw. Inhalte werden während des Transportweges durch Verschlüsselungsmaßnahmen unkenntlich gemacht und liegen teilweise in verschlüsselter Form in Datenbanken einer Organisation. Es lassen sich längst Entwicklungstrends feststellen, die das Bereitstellen von Web Sites ermöglichen, die Inhalte ausschließlich in verschlüsselter Form empfangen, speichern und ausliefern. Lediglich der einzelne Nutzer kann seine eigenen Inhalte nutzbar machen und interpretieren. Eine flächendeckende Adaption dieser Systemlösungen ist jedoch noch nicht feststellbar.<sup>213</sup>

Neben dem konsequenten Abschirmen aus Sicherheitsgründen werden einige IT-Systeme bewusst über Schnittstellen miteinander verbunden.<sup>214</sup> Typische Vertreter sind z. B. Web-Site-Analyse-Tools, die Auswertungen von Benutzeraktivitäten ermöglichen und unmittelbar mit der Web Site verbunden sind.<sup>215</sup> Darüber hinaus ist eine Interoperabilität mit Enterprise-Ressource-Planning-Systemen (ERP-Systemen) oder Customer-Relationship-Management-Systemen (CRM-Systemen) beobachtbare Praxis. Die Web Site ist ein technischer Bestandteil der Wertschöpfung und wird zum Informationsaustausch zwischen diesen IT-Systemen konfiguriert.<sup>216</sup>

---

211 Vgl. Sackmann, Steffan: IT-Sicherheit, Online im Internet: <http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/technologien-methoden/Informatik--Grundlagen/IT-Sicherheit/index.html>, 01.03.2018.

212 Vgl. Nahari, Hadi; Krutz, Ronald L.: Web commerce security: design and development, Indianapolis, Ind: Wiley 2011, S. 271 ff. oder Sackmann, Steffan: IT-Sicherheit, a. a. O., 01.03.2018.

213 Vgl. o. V.: Blockchain – Meine Daten gehören mir, in: wisu - Das Wirtschaftsstudium, 4/2018, S. 344 ff.

214 Vgl. zur „Interoperabilität“ Bergsmann, Johannes; Unterauer, Markus: Requirements Engineering für die agile Softwareentwicklung: Methoden, Techniken und Strategien, 1. Aufl, Heidelberg: dpunkt-Verl 2014, S. 110.

215 Vgl. Hassler, Marco: Web Analytics: Metriken auswerten, Besucherverhalten verstehen, Website optimieren, a. a. O., S. 40.

216 Vgl. Hillebrand, Ralf-Thomas: Online-Kommunikation für Verbände: wie Ihre Botschaften die Zielgruppen sicher erreichen und überzeugen, Wiesbaden: Springer Gabler 2018, S. 137 f.

### 3.5 Zusammenfassung und Erläuterung des Fortgangs der Arbeit

Die Eigenschaften einer Web Site wurden in inhaltliche, gestalterische und betriebsspezifische Eigenschaften gegliedert. Systemtheoretisch stellt diese Gliederung ein vollständiges Strukturmodell der Eigenschaften einer Web Site dar.<sup>217</sup>

Die geschilderte Menge der Eigenschaften in diesem Modell ist auf das Konstruktionsziel der vorliegenden Arbeit zugeschnitten. Das Modell und die Eigenschaften basieren auf Praxisbeobachtungen und der Auswertung der einschlägigen Fachliteratur.

Aufbauend auf der Eigenschaftsanalyse folgt im anschließenden Kapitel eine Analyse der Anforderungen an ein IT-System zum Erfassen, Bewahren und Bereitstellen einer WSB. Die Anforderungen an ein solches Web-Site-Biographie-System (WSBS) bestimmen funktionale und nicht-funktionale Aspekte der Systemlösung.<sup>218</sup>

---

217 Vgl. Zschau, O.; Traub, D.; Zahradka, R.: Web Content Management. Websites professionell planen und betreiben. 2., überarb. und erw. Aufl, a. a. O., S. 59 f. oder Krüger, Jörg Dennis; Kopp, Matthias: Web-Content managen: professioneller Einsatz von Content-Management-Systemen, a. a. O., S. 169.

218 Vgl. Balzert, Helmut: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, 3. Aufl, Heidelberg: Spektrum, Akad. Verl 2009, S. 488 f.

## 4 Anforderungsanalyse eines WSBS

### 4.1 Systematisierung der Anforderungsanalyse

Die Gliederung der Anforderungsanalyse folgt dem klassischen Dreiklang fachlicher, organisatorischer und technischer Anforderungen an IT-Systeme. Die fachlichen Anforderungen beschreiben den funktionalen Umfang eines IT-Systems, technische und organisatorische Anforderungen beschreiben dagegen die Infrastruktur-Komponenten.<sup>219</sup> Die Aufteilung der Anforderungen in eine fachliche, organisatorische und technische Dimension orientiert sich an der ganzheitlichen Betrachtungsweise der Systementwicklung.<sup>220</sup>

Damit in der vorliegenden Arbeit eine systematische Analyse der Anforderungen durchgeführt werden kann, wurden signifikante Eigenschaften einer Web Site definiert. Diese Eigenschaften nehmen unmittelbar Einfluss auf das Konstruktionsziel der Arbeit und bilden das Fundament der Anforderungsanalyse.<sup>221</sup> Nur wenn die WSB alle signifikanten Eigenschaften einer Web Site reproduzieren kann, ist das Erlebbarmachen der digitalen Vergangenheit einer Organisation möglich.<sup>222</sup>

Die Summe der Anforderungen beschreibt den Sollzustand des zu realisierenden Zielsystems.<sup>223</sup> Das nachfolgende Kapitel 4.2 widmet sich zunächst der fachlichen Anforderungsanalyse.

---

219 Vgl. Balzert, Helmut: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, a. a. O., S. 456.

220 Vgl. Balzert, Helmut: Lehrbuch der Software-Technik, a. a. O., S. 24 ff. oder auch Bergsmann, Johannes; Unterauer, Markus: Requirements Engineering für die agile Softwareentwicklung: Methoden, Techniken und Strategien, a. a. O., S. 97 ff.

221 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 56.

222 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 56 und auch Rosenthal, David; Robertson, Thomas; Lipkis, Tom; Reich, Vicky; Morabito, Seth: Requirements for Digital Preservation Systems – A Bottom-Up Approach, a. a. O., 12.08.2017.

223 Vgl. Balzert, Helmut: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, a. a. O., S. 456.

## 4.2 Fachliche Anforderungsanalyse

### 4.2.1 Systematisierung der fachlichen Anforderungsanalyse

Das Ziel der fachlichen Anforderungsanalyse ist das Formalisieren von funktionalen Gegebenheiten, die das Realisieren einer WSB ermöglichen.<sup>224</sup> Die Summe der fachlichen Anforderungen beschreibt den Leistungsumfang eines IT-Systems und ermöglicht die systematische Konzeption der Systemlösung.<sup>225</sup>

Abbildung 14 stellt die Systematik der fachlichen Anforderungsanalyse beispielhaft dar.

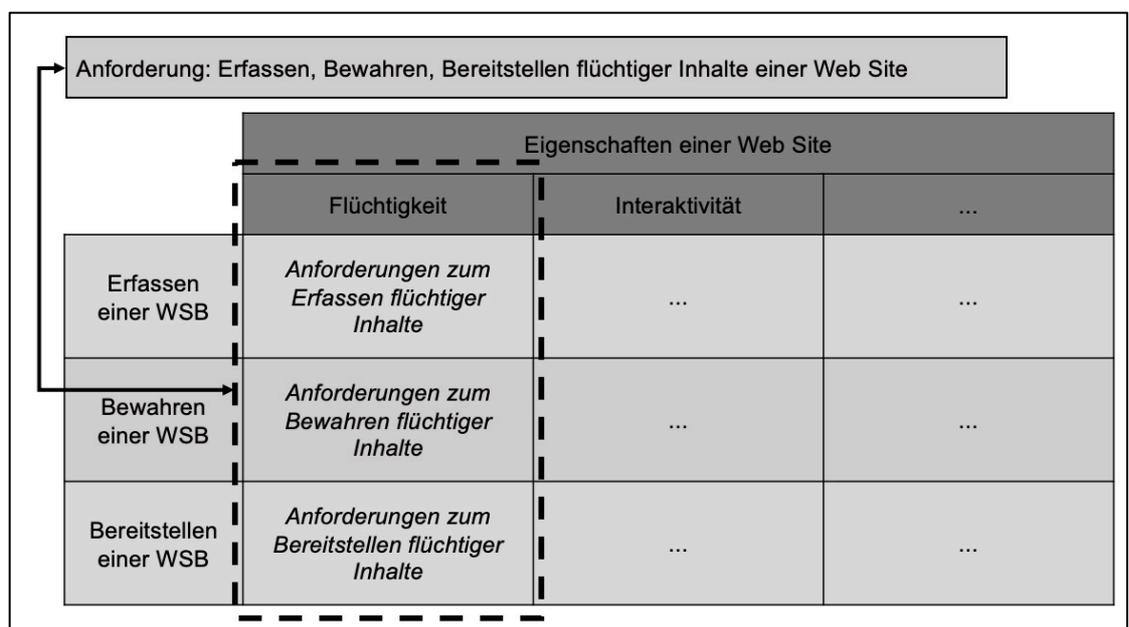


Abb. 14: Systematisierung der fachlichen Anforderungen

Die fachlichen Anforderungen werden auf Basis der Web-Site-Eigenschaften aus Kapitel 3 strukturiert und durch Aufgaben untergliedert. Die durch das IT-System zu erfüllenden Aufgaben sind das Erfassen, Bewahren und Bereitstellen einer WSB.

224 Vgl. Ebert, Christof: Systematisches Requirements Engineering: Anforderungen ermitteln, spezifizieren, analysieren und verwalten, 5., überarb. Aufl., Heidelberg: dpunkt-Verl 2014, S. 133.

225 Vgl. Ludwig, Jochen; Lichter, Horst: Software Engineering: Grundlagen, Menschen, Prozesse, Techniken, a. a. O., S. 369 ff.

#### 4.2.2 Anforderungen auf Basis inhaltlicher Eigenschaften

Die Zustände einer Web Site sollen auf unbestimmte Zeit erhalten bleiben und für die Nachwelt erlebbar gemacht werden. Um dieses Ziel zu erreichen, müssen Inhalte einer Web Site in einem digitalen Abbild erfasst, bewahrt und bereitgestellt werden.<sup>226</sup> Im Folgenden werden Anforderungen an diese Aufgaben formuliert.

##### *Erfassen, Bewahren, Bereitstellen flüchtiger Inhalte einer Web Site*

Es ist eine allgegenwärtige und in der Praxis beobachtbare Tatsache, dass Inhalte auf einer Web Site einem Wandel unterliegen.<sup>227</sup> Systemfunktionen zur Erfassung einer WSB müssen daher ereignisgesteuert und periodisch das Kopieren dieser Inhalte ermöglichen. Erfassungsfrequenzen und -prozesse sollen systembasiert geplant und ausgeführt werden können. Stark flüchtige Inhalte müssen häufiger abgefragt und in Abbilder überführt werden als langfristig beständige Inhalte. Auf diese Weise lassen sich nicht nur vollständige Kopien einer Web Site erstellen, sondern auch der Einsatz von IT-Ressourcen (Rechenleistung etc.) optimieren.

Das Generieren der Abbilder ist idealerweise isoliert vom Analysieren der Inhaltsänderungen durchzuführen. Beim Kopieren großer Datenmengen kann das Erfassen eines Abbildes eine vergleichsweise lange Zeit in Anspruch nehmen. Durch das Entkoppeln der Analyse- und Kopierfunktion soll verhindert werden, dass während der Erfassung Inhaltsänderungen verpasst und nicht als Bestandteil der WSB geführt werden.

Innerhalb der Bewahrungsumgebung muss eine zeitliche Einordnung sämtlicher Abbilder erfolgen. Jede gesicherte Ressource muss nicht nur ein eindeutiges Identifizierungsmerkmal erhalten, sondern auch einen Zeitstempel als Dokumentation des Erstellungszeitpunktes besitzen. Dieser Erstellungszeitpunkt soll zum einen die systeminterne Verwaltung unterstützen, zum anderen aber auch zum Auffindbarmachen und Interpretieren der

---

226 Vgl. in diesem Zusammenhang Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 56 f.

227 Vgl. Brown, Adrian: Archiving websites: a practical guide for information management professionals, a. a. O., S. 34 f. oder Hale, Scott A.; Blank, Grant; Alexander, Victoria D.: Live versus archive: Comparing a web archive to a population of web pages, a. a. O., S. 45 f. oder Brügger, Niels: Web History, an Emerging Field of Study, a. a. O., S. 7 f. und Adar, Eytan; Teevan, Jaime; Dumais, Susan T.; Elsas, Jonathan L.: The web changes everything: understanding the dynamics of web content, a. a. O., S. 284 ff.

Abbilder dienen.<sup>228</sup> Durch das Präsentieren und Suchen einzelner Abbilder soll Transparenz und Recherchierbarkeit ermöglicht werden.<sup>229</sup>

*Erfassen, Bewahren, Bereitstellen interaktiver Inhalte einer Web Site*

Neben der Flüchtigkeit von Inhalten, ist eine Web Site durch Interaktivität geprägt.<sup>230</sup> Inhalte werden mit Programmkomponenten angereichert, die auf Eingaben der Nutzer reagieren. Für die Erstellung einer WSB sind diese Inhalte abzubilden.

Während der Erfassung sollen alle Inhalte identifiziert und verarbeitet werden, die eine Interaktion mit Nutzern der Web Site voraussetzen. Skript-Dateien und Programme zur Realisierung interaktiver Inhalte sind zu kopieren und als Bestandteil der Abbilder einer Web Site zu führen. Kopien der Skript-Dateien und der Programme werden zur Bewahrung aufbereitet und müssen durch Erhaltungsmechanismen langfristig konserviert werden. Beispielhaft ist die Durchführung von Migrationsverfahren zu nennen.

Zur Bereitstellung interaktiver Inhalte müssen die kopierten Skript-Dateien und Programme ausgeliefert und im Web Browser eines Nutzers erlebbar gemacht werden. Für veraltete Dateiformate, die nicht migriert werden können, sollte eine Emulationsfunktion das Reproduzieren in angepassten, technischen Umgebungen ermöglichen.<sup>231</sup>

*Erfassen, Bewahren, Bereitstellen abgesicherter Inhalte einer Web Site*

Neben öffentlichen bzw. frei zugänglichen Inhalten existieren Inhalte auf einer Web Site, die nur nach erfolgreicher Authentifizierung und Sicherheitsabfragen für einzelne Nutzerprofile abgerufen werden können.<sup>232</sup> Die geforderte Systemlösung muss daher Funktionen aufweisen, die einen Zugriff auf diese abgesicherten Inhalte ermöglichen.<sup>233</sup> Im Fokus der Funktionen liegt eine historische Dokumentation der Inhalte, nicht die rechts-

---

228 Vgl. InSPECT-Katalog in Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 34.

229 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 65 f.

230 Vgl. Masanès, Julien: Web Archiving: Issues and Methods, a. a. O., S. 14 ff. oder Kappel, Gerti; Pröll, Birgit; Reich, Siegfried; Retschitzegger, Werner: Web Engineering – Die Disziplin zur systematischen Entwicklung von Web-Anwendungen, a. a. O., S. 7. oder Krause, Dirk: VR im Web: Virtual-Reality-Web-Apps mit A-Frame programmieren, a. a. O., S. 116 ff.

231 Vgl. Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften, DARIAH-DE (Hrsg.): Handbuch Digital Humanities: Anwendungen, Forschungsdaten und Projekte / DARIAH-DE, a. a. O., S. 65.

232 Vgl. Mitchell, Ryan E.: Web scraping with Python: collecting data from the modern web, a. a. O., S. 208 ff.

233 Vgl. Masanès, Julien: Web archiving, a. a. O., S. 104.

sichere Erfassung sämtlicher Nutzeraktivitäten. Inhalte sollen auf Basis eines nicht-personalisierten Demo-Kontos abgerufen, analysiert und kopiert werden. Auf diese Weise werden rechtliche Problemstellungen, die durch eine Erfassung von realen Nutzeraktivitäten induziert werden, exkludiert.

Zur Bewahrung der abgesicherten Inhalte müssen Berechtigungen konfigurierbar und verwaltbar sein. Zugriffsrechte stellen im Sinne eines administrativen Kontextes sicher, dass ausgewählte Abbilder nur nach einer erfolgreichen Authentifizierung administrierbar oder erlebbar sind. Es soll verhindert werden, dass sensible Inhalte uneingeschränkt zugänglich sind.<sup>234</sup> Durch eine geeignete Nutzerverwaltung und Berechtigungsgruppen lassen sich granulare Berechtigungen vergeben und Restriktionen für das Abrufen der Abbilder implementieren.

Die Bereitstellung abgesicherter Inhalte sollte durch Funktionen umgesetzt sein, die eine Authentifizierung voraussetzen. Das gesicherte Nutzerprofil zu einem Demo-Konto steht beispielsweise erst nach der Eingabe geeigneter Zugangsdaten zum Abrufen bereit. Einzelne Dateien, die auf der originalen Web Site mit einem Passwort geschützt werden, sind in der Bereitstellungsumgebung weiterhin durch das gleiche Passwort geschützt.

#### *Erfassen, Bewahren, Bereitstellen kritischer Inhalte einer Web Site*

Kritische Inhalte unterliegen Regelungen und Sanktionen seitens des Gesetzgebers. Neben Persönlichkeitsrechten, die mit der Veröffentlichung von Mitarbeiterfotos einhergehen, gilt es unter anderem, das Urheberrecht zu wahren.<sup>235</sup> Darüber hinaus sind auch Regelungen ausschlaggebend, die sich aus dem Datenschutz, Verwertungsrechten oder der Compliance einer Organisation ableiten lassen.<sup>236</sup>

Staatliche und interne Regelwerke müssen nicht nur während der Veröffentlichung von Inhalten auf einer Web Site, sondern auch im Zusammenhang mit der langfristigen Bewahrung berücksichtigt werden. Wenn Inhalte auf Grund von gesetzlichen Regelungen von der originalen Web Site entfernt werden, sollte dies auch in der WSB geschehen.<sup>237</sup>

Die Organisation muss daher die Möglichkeit haben, proaktiv und reaktiv Inhalte einzelner Abbilder auszuschließen bzw. zu entfernen. Sind kritische Inhalte bereits im Vorfeld eines Erfassungsprozesses bekannt, sollen diese proaktiv von der Erfassung exkludiert

---

234 Vgl. zu administrativen Metadaten Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 68 f.

235 Vgl. in diesem Zusammenhang dejure.org Rechtsinformationssysteme GmbH (Hrsg.): Rechtsprechung LAG Hessen, 24.01.2012 - 19 SaGa 1480/11, a. a. O., 03.02.2018.

236 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 51.

237 Vgl. z. B. Brown, Adrian: Archiving websites: a practical guide for information management professionals, a. a. O., S. 147.

werden können. Administratoren sollen die Möglichkeit haben, eine Negativliste zu führen, die festlegt, welche Inhalte bei der Erfassung ausgeschlossen werden.

Darüber hinaus muss eine Manipulationsfunktion vorhanden sein, die Inhalte auch lange nach ihrer Erfassung auffindbar und editierbar macht. Die Manipulationen der Abbilder sind zur Wahrung der Authentizität in Metadaten zu dokumentieren. Auf diese Weise wird eine Nachvollziehbarkeit und Integrität des Datenbestands gewährleistet.<sup>238</sup>

#### *Erfassen, Bewahren, Bereitstellen personalisierter Inhalte einer Web Site*

Auf vielen Web Sites ist das Bereitstellen von personalisierten Inhalten gegenwärtig. Empfehlungssysteme speichern die Eingaben der Nutzer, analysieren diese und generieren individuellen Output.<sup>239</sup>

Besonders vor dem Hintergrund rechtlicher Regelwerke ist das Erfassen, Bewahren und Bereitstellen dieser personalisierten Inhalte kritisch zu bewerten.<sup>240</sup> Der Aufwand zur Legitimation einer Speicherumgebung, die personalisierte Inhalte zur weiteren Verwendung aufbewahrt, ist durchaus erheblich bzw. eine Legitimation entsprechender Systemumgebungen schwer umsetzbar.<sup>241</sup> Darüber hinaus vermitteln personalisierte Inhalte alleinstehend wenig Erkenntnis über die Zustände der digitalen Vergangenheit im Web, da häufig nur kleine Inhaltsbereiche oder Gestaltungselemente an den Nutzerkontext einer Web Site angepasst werden. Es soll daher in der vorliegenden Arbeit von der Erfassung, Bewahrung und Bereitstellung personalisierter Inhalte Abstand genommen werden.

Nachdem im vorliegenden Kapitel die fachlichen Anforderungen auf Basis der inhaltlichen Eigenschaften formuliert wurden, folgt im anschließenden Kapitel die Anforderungserhebung auf Basis der gestalterischen Eigenschaften einer Web Site.

---

238 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 60.

239 Vgl. Klahold, André: Empfehlungssysteme: Recommender Systems – Grundlagen, Konzepte und Lösungen, a. a. O., S. 4 ff.

240 Vgl. Art. 6 EU-DSGVO im Amtsblatt der Europäischen Union (Hrsg.): VERORDNUNG (EU) 2016/679 DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG, a. a. O., 05.02.2018.

241 Vgl. Fitch, K.: Web site archiving: an approach to recording every materially different response produced by a Website, a. a. O., S. 6 ff.

### 4.2.3 Anforderungen auf Basis gestalterischer Eigenschaften

Damit originalgetreue Abbilder einer Web Site entstehen, müssen nicht nur inhaltliche Eigenschaften, sondern auch gestalterische Eigenschaften in der WSB reproduziert werden.<sup>242</sup> In Kapitel 3 wurden vier gestalterische Eigenschaften definiert. Aufbauend auf diesen vier gestalterischen Eigenschaften wird im folgenden Abschnitt eine fachliche Anforderungsanalyse durchgeführt.

#### *Erfassen, Bewahren, Bereitstellen der grafischen Aufbereitung einer Web Site*

Zur individuellen Präsentation der Organisation im digitalen Raum werden die Inhalte auf einer Web Site klassischerweise grafisch aufbereitet.<sup>243</sup> Es lassen sich Farben für Texte, Links oder Umrandungen definieren, Hintergründe gestalten und individuelle Schriftarten bestimmen. Inhalte können hervorgehoben werden, um die Aufmerksamkeit der Nutzer zu wecken.<sup>244</sup> Zudem lässt sich die Web Site im Kontext des allgemeinen Erscheinungsbildes einer Organisation präsentieren und die gestalterische Identität des Unternehmens in den digitalen Raum transportieren.<sup>245</sup> Dem Grundsatz einer konsistenten Trennung von Inhalt und Gestaltung einer Web Site folgend, werden die Befehle zur grafischen Aufbereitung im Regelfall in Style Sheets ausgelagert.<sup>246</sup>

Style Sheets müssen im Erfassungsprozess identifiziert, zur Bewahrung kopiert und mit relevanten Abbildern verknüpft werden. Darüber hinaus sind auf vergleichbare Weise alle grafischen Determinanten zu sichern, die von externen Services oder Dienstleistern stammen. Hier gilt es z. B. Schriftarten zu kopieren, die durch das Einbinden von dafür vorgesehenen Dateiformaten charakterisiert sind.<sup>247</sup> Zur Bereitstellung sind alle gestaltgebenden Dateien in einem Abbild zusammenzuführen, um die Reproduktion des Originals bestmöglich zu unterstützen.

---

242 Vgl. in diesem Zusammenhang Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 56 f.

243 Vgl. W3C - World Wide Web Consortium (Hrsg.): WHAT IS CSS?, a. a. O., 10.02.2018.

244 Vgl. Meidl, Oliver: Global Website: Webdesign im internationalen Umfeld, a. a. O., S. 22.

245 Vgl. Meidl, Oliver: Global Website: Webdesign im internationalen Umfeld, a. a. O., S. 19.

246 Vgl. Bühler, Peter; Patrick Schlaich; Sinner, Dominik: HTML5 und CSS3: Semantik - Design - Responsive Layouts, a. a. O., S. 45.

247 Vgl. Bühler, Peter; Patrick Schlaich; Sinner, Dominik: HTML5 und CSS3: Semantik - Design - Responsive Layouts, a. a. O., S. 60.

*Erfassen, Bewahren, Bereitstellen der Positionierungen auf einer Web Site*

Web Site Designer können bei der Positionierung von Inhalten zwischen fluiden, adaptiven und reagierenden Positionierungen wählen.<sup>248</sup> Während fluide Positionierungen durch eine prozentuale Aufteilung umgesetzt werden, verändern adaptive Positionierungen die Anzahl und Menge der Darstellungselemente. Je nach Bildschirmgröße wird eine passende Anordnung von Grafiken und Inhalten vorgenommen. Die reagierende Positionierung vereint dagegen beide Konzepte.<sup>249</sup>

Ein WSBS soll in der Lage sein, Positionierungen der Inhalte und Grafiken auf einer Web Site originalgetreu zu reproduzieren. Beim Erfassen müssen sämtliche Befehle kopiert werden, die für die Positionierung verantwortlich sind. Bei reagierenden oder adaptiven Positionierungen werden Skripte oder einzelne Gestaltungsbefehle je nach Auflösung individuell nachgeladen.<sup>250</sup> Für die vollständige Erfassung der Gestaltungsformen müssen diese endgeräteabhängigen Positionierungen berücksichtigt werden.

In der WSB sollen die unterschiedlichen Versionen der Web Site inkludiert werden. Nutzer sollen die Möglichkeit haben, zwischen verschiedenen Gestaltungsformen zu wählen, falls diese vorhanden sind bzw. erfasst wurden.

*Erfassen, Bewahren, Bereitstellen der Web-Site-Strukturierung*

Aus einer technisch-konstruktiven Sicht ist die Web Site als ein sach-logischer Verbund von HTML-Dateien charakterisiert, die über Links miteinander verbunden sind.<sup>251</sup> Durch die systematische Verbindung von HTML-Dateien werden Navigationsstrukturen geschaffen, die das technische Konstrukt „Web Site“ bilden.<sup>252</sup> Die WSB soll sämtliche HTML-Dateien bzw. Web-Seiten im Adressraum einer Organisation abbilden. Ausgehend von einer Startseite werden alle Web-Seiten der Web Site aufgerufen, analysiert und gesichert. Während der Erfassung müssen geeignete Funktionen sämtliche Verknüpfungen zwischen den Web-Seiten erfassen und entlang dieser virtuellen Pfade Abbilder einzelner Web-Seiten generieren.<sup>253</sup> Die Verknüpfungen zwischen einzelnen Web-Seiten

---

248 Vgl. auch hier Bühler, Peter; Schlaich, Patrick; Sinner, Dominik: Webdesign: Interfacedesign - Screendesign - Mobiles Webdesign, a. a. O., S. 47.

249 Vgl. Carlos, Gilberto: Responsive Web Design with jQuery, a. a. O., S. 12 ff.

250 Vgl. Bühler, Peter; Schlaich, Patrick; Sinner, Dominik: Webdesign: Interfacedesign - Screendesign - Mobiles Webdesign, a. a. O., S. 50.

251 Vgl. Schwickert, Axel C.: Web Site Engineering: Ökonomische Analyse und Entwicklungssystematik für eBusiness-Präsenzen, a. a. O., S. 98 ff.

252 Vgl. Bühler, Peter; Schlaich, Patrick; Sinner, Dominik: Webdesign: Interfacedesign - Screendesign - Mobiles Webdesign, a. a. O., S. 26 ff.

253 Vgl. Pant, Gautam; Srinivasan, Padmini; Menczer, Filippo: Crawling the Web, a. a. O., S. 153. oder Alpar, Andre; Koczy, Markus; Metzen, Maik: SEO - Strategie, Taktik und Technik: Online-Marketing mittels effektiver Suchmaschinenoptimierung, a. a. O., S. 47 f.

sollen in der Bewahrungsumgebung dokumentiert und zum Erleben reproduziert werden. Die Links zwischen Web-Seiten werden als Bestandteil der WSB gespeichert und in den Abbildern nutzbar gemacht.

Es gilt in diesem Zusammenhang zu berücksichtigen, dass Unternehmen bzw. Organisationen auch Links verwenden, die auf Web Sites fremder Organisationen verweisen, z. B. im Rahmen einer Werbemaßnahme oder zum Darstellen von Kooperationspartnern. Eine Funktionsmenge muss also sicherstellen, dass nur Inhalte unter zuvor festgelegten und legitimierten Web-Adressen analysiert und abgebildet werden. Web-Adressen, die nicht in den Verantwortungsbereich der eigenen Organisation fallen, sollen in einer Positivliste gepflegt werden. Ist dies nicht der Fall, kann bei der Erfassung urheberrechtlich geschützter Inhalt kopiert werden. Mit Blick auf rechtliche Konsequenzen gilt es diese Situation kontrollierbar zu machen.

#### *Erfassen, Bewahren, Bereitstellen der Vielgestaltigkeit einer Web Site*

Bei der Gestaltung einer Web Site ist es nicht unüblich, dass die grafische Aufbereitung im Zeitablauf wechselt und neue Positionierungen implementiert werden.<sup>254</sup> Ein WSBS soll daher während der Erfassungsprozesse nicht nur Inhalte auf Veränderungen überprüfen, sondern auch sämtliche Ressourcen der Gestaltung mit einbeziehen.

Neben der Anpassung an Endgeräte oder Bildschirmauflösungen sind saisonale Anpassungen der Gestaltung allgegenwärtig. Es findet bei einigen Web Sites ein periodischer Wechsel der grafischen Aufbereitung und Positionierung statt. Auslöser dieser saisonalen Anpassungen sind vielfältig.<sup>255</sup>

Im Rahmen der Erfassung sollen sämtliche Veränderungen in der Gestaltung dokumentiert werden. Für ein zielkonformes System sind der HTML-Quellcode und die darin verlinkten Gestaltungsressourcen (Skripte, Style Sheets etc.) kontinuierlich auf Veränderungen zu überprüfen. Festgestellte Änderungen sind abzubilden, zu bewahren und als Bestandteil der WSB bereitzustellen.

Werden Veränderungen in der grafischen Aufbereitung oder in der Positionierung einer Web Site festgestellt, sind diese unmittelbar in Abbildern festzuhalten. Die einzelnen Abbilder werden in der Bewahrungsumgebung mit Identifikationsmerkmalen bzw. Zeitstempeln versehen und für die Nachwelt konserviert.

---

254 Vgl. Galitz, Wilbert O.: The essential guide to user interface design: an introduction to GUI design principles and techniques, a. a. O., S. 258 ff.

255 Vgl. Sens, Bastian: Schluss mit 08/15-Websites – so bringen Sie Ihr Online-Marketing auf Erfolgskurs: 79 Tipps für Ihren Online-Auftritt, a. a. O., S. 14 ff.

Im Rahmen der Bereitstellung sind alle gestaltgebenden Dateien und Befehle als Bestandteil der Abbilder erlebbar zu machen.

Nachdem im vorliegenden Kapitel die fachlichen Anforderungen auf Basis der gestalterischen Eigenschaften einer Web Site formuliert wurden, folgt abschließend eine Anforderungserhebung auf Basis der betriebsspezifischen Eigenschaften.

#### 4.2.4 Anforderungen auf Basis betriebsspezifischer Eigenschaften

Inhalt und Gestalt einer Web Site sind das Ergebnis technischer und organisatorischer Operationen, die auf Basis der Infrastruktur einer Organisation durchgeführt werden.<sup>256</sup>

Um möglichst vollständige Abbilder einer Web Site zu erzeugen, ist die Dokumentation der technisch-organisatorischen Entstehung zu inkludieren.<sup>257</sup> Für die Entwicklung eines entsprechenden Funktionsumfangs werden im vorliegenden Kapitel geeignete Anforderungen auf Basis betriebsspezifischer Eigenschaften einer Web Site definiert.

##### *Erfassen, Bewahren, Bereitstellen der Formatvielfalt einer Web Site*

Dateien werden auf einer Web Site in unterschiedlichen Dateiformaten publiziert.<sup>258</sup> Charakteristisch für die Ausprägungen der Dateiformate ist eine individuelle Struktur der Datenströme ebenso wie divergierende und teilweise proprietäre Software, die zum Erstellen und Interpretieren verwendet wird.<sup>259</sup>

Für eine vollständige Reproduktion der WSB ist es notwendig, dass eingesetzte Dateien und ihre Dateiformate abgefragt und überprüft werden können.<sup>260</sup> Die Dateien und Dateiformate müssen fehlerfrei verarbeitet werden. Integrität und Authentizität sind relevante Prüfungsaspekte, die zur Validierung der Dateien herangezogen werden sollen.<sup>261</sup>

---

256 Vgl. zur Definition der IT-Infrastruktur Weill, Peter; Broadbent, Marianne: Leveraging the new infrastructure: how market leaders capitalize on information technology, Boston, Mass: Harvard Business School Press 1998, S. 86.

257 Vgl. Aschenbrenner, Andreas; Rauber, Andreas: Mining Web Collections, in: Web Archiving, Hrsg.: Masanès, Julien, Berlin, Heidelberg: Springer-Verlag 2010, S. 159 f.

258 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 12 f. oder Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften, DARIAH-DE (Hrsg.): Handbuch Digital Humanities: Anwendungen, Forschungsdaten und Projekte / DARIAH-DE, a. a. O., S. 68 f.

259 Vgl. z. B. Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 16 f.

260 Vgl. Hansch, Pierre; Rentschler, Christian: EmotionWeb: emotionale Websites durch Bewegtbild und Sound-Design, a. a. O., S. 19 ff.

261 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 57 ff.

Um die Sicherheit des IT-Systems zu wahren, muss eine Liste akzeptierter Dateiformate konfigurierbar sein. Die WSB bildet nur Dateien ab, deren Dateiformate innerhalb dieser Liste geführt und damit wissentlich akzeptiert werden.<sup>262</sup>

Zur Reproduktion und Interpretation der Dateiformate müssen technische Merkmalsbeschreibungen einer Datei in der Bewahrungsumgebung dokumentiert werden. Darüber hinaus sind standardisierte deskriptive Attribute wie ein Identifizierungsmerkmal, der Name der Datei und ein Erfassungszeitpunkt zu hinterlegen. Die Ausprägung der zu erfassenden Merkmale soll auf Basis etablierter und standardisierter Rahmenwerke erfolgen, die entsprechende Merkmale bereits für die digitale Erhaltung definieren.<sup>263</sup>

Neben Merkmalsbeschreibungen müssen Abbilder einer Web Site als Konglomerate heterogener Dateiformate gespeichert werden. Alle mit einer Web Site verknüpften Dateien sind in einer strukturierten Speicherlösung zu verwahren und für zukünftige Replikationen auffindbar zu machen.<sup>264</sup>

Eine WSB stellt das Erlebbarmachen der Vergangenheit in den Fokus und soll eine möglichst performante Bereitstellung der Abbilder ermöglichen. Aus diesem Grund wird eine leistungsstarke Speicherlösung erforderlich, die Abbilder für die Nachwelt konserviert und abrufbar macht.<sup>265</sup>

Die Bereitstellung der Speicherlösung soll flexibel an neue Web-Technologien angepasst werden können, um jederzeit zeitgemäße Bedienoberflächen mit geringem Anpassungsaufwand zu erstellen.<sup>266</sup>

Flexibilität soll durch eine Schnittstelle entstehen, die sämtliche Informationen und Inhalte eines Abbildes in einem standardisierten Format ausliefert. Die Interpretation bzw. Zusammenstellung eines Abbildes erfolgt im Web Browser des Nutzers.<sup>267</sup>

---

262 Vgl. Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 40 ff.

263 Vgl. Schimpf, Sabine: Das OAIS-Modell für die Langzeitarchivierung: Anwendung der ISO 14721 in Bibliotheken und Archiven, a. a. O., S. 128.

264 Vgl. Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 44.

265 Vgl. Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 58 ff.

266 Vgl. Masanès, Julien: Web archiving, a. a. O., S. 32 ff.

267 Vgl. Patni, Sanjay: Pro RESTful APIs: Design, Build and Integrate with REST, JSON, XML and JAX-RS, Berkeley, CA: Apress Imprint : Apress 2017, S. 66 ff.

*Erfassen, Bewahren, Bereitstellen der Infrastruktur einer Web Site*

Wie bereits in der Eigenschaftsanalyse beschrieben, lässt sich die Infrastruktur einer Web Site in eine technische und organisatorische (IT-)Infrastruktur gliedern.<sup>268</sup> Für die vollständige Dokumentation einer WSB müssen die technischen Komponenten zum Betrieb einer Web Site ebenso dokumentiert werden wie die institutionellen bzw. organisatorischen Merkmale.<sup>269</sup> Das WSBS soll daher das Erfassen, Bewahren und Bereitstellen von *Kontext-Informationen* zum Beschreiben der Infrastruktur ermöglichen.

Kontext-Informationen beschreiben die Abbilder einer Web Site und beinhalten Informationen über die organisatorische und technische Betriebsumgebung. Damit eine einheitliche Erfassung der Kontext-Informationen erfolgen kann, muss deren Struktur vordefiniert werden. Kategorien sollen eine Klassifikation und Ordnung der Kontext-Informationen ermöglichen. Mit einem Set an Kategorien soll vorgegeben werden, dass z. B. Informationen zum Server-Hersteller, zur Prozessorleistung oder zum eingesetzten Betriebssystem erfasst werden müssen, um die technische IT-Infrastruktur einer Web Site vollständig zu dokumentieren. Kategorien und Kontext-Informationen lassen sich nachträglich erstellen, verändern oder löschen.

Vergleichbar mit einem archivierten Foto, welches klassischerweise nicht nur mit Informationen über die Kamera, sondern auch mit Informationen über den Fotografen dokumentiert wird, sollen für Abbilder einer Web Site auch organisatorische Informationen zur Infrastruktur hinterlegt werden.<sup>270</sup> Beispielhaft sind an dieser Stelle Kontext-Informationen anzuführen, die Aufschluss über Personen geben, die mit der Erstellung von Inhalten der Web Site betraut waren.<sup>271</sup>

---

268 Vgl. Weill, Peter; Broadbent, Marianne: Leveraging the new infrastructure: how market leaders capitalize on information technology, a. a. O., S. 86 oder Rudolph, Simone: Servicebasierte Planung und Steuerung der IT-Infrastruktur im Mittelstand: ein Modellansatz zur Struktur der IT-Leistungserbringung, a. a. O., S. 13 ff. oder auch Office of Government Commerce (OGC) (Hrsg.): The official introduction to the ITIL service lifecycle, a. a. O., S. 199.

269 Vgl. auch Rumpel, Rainer J.: Planung und Realisierung von IT-Infrastrukturen - ein prozessbasierter Ansatz, München: Oldenbourg 2012, S. 11 f. und Weill, Peter; Broadbent, Marianne: Leveraging the new infrastructure: how market leaders capitalize on information technology, a. a. O., S. 86.

270 Vgl. Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 36.

271 Vgl. auch hier Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 36.

### *Berücksichtigung der Verteilung einer Web Site*

Um möglichst geringe Ladezeiten, wenige Ausfälle und hohe Verfügbarkeit zu gewährleisten, sind einige Web Sites nicht nur durch eine Verteilung auf mehrere Server ausgezeichnet, sondern auch durch externe Bilddatenbanken oder Programmbibliotheken.<sup>272</sup>

Während der Erfassung müssen nicht nur organisationseigene Ressourcen zum Darstellen einer Web Site kopiert und in die Bewahrungsumgebung überführt werden. Unter Umständen sind auch Dateien von Servern fremder Institutionen zu kopieren. Um rechtliche Konsequenzen zu kontrollieren, soll eine Positivliste geführt werden, die dokumentiert, auf welche Web-Adressen externer Dienstleister zugegriffen werden darf. Nur auf diese Weise lässt sich eine kontrollierte Verfügungsgewalt über externe Inhalte und Gestaltungselemente gewährleisten und die Beständigkeit einer WSB sicherstellen.<sup>273</sup>

Im Rahmen der Bereitstellung sollen Abbilder ausschließlich Inhalte und Gestaltungselemente verwenden, die für eine WSB kopiert und in die Bewahrungsumgebung überführt wurden. Das Einbinden und Reproduzieren von Dateien, die nicht in den Datenbanken der Organisation gespeichert sind, muss unterbunden werden. Nur so ist eine vollständige Sicherung der Abbilder im Verfügungsraum der Organisation möglich.<sup>274</sup>

### *Berücksichtigung der Sicherheit einer Web Site*

Zur Wahrung der Sicherheit soll auf funktionaler Ebene gewährleistet sein, dass das Erfassen der WSB keine negativen Auswirkungen auf den produktiven Betrieb der originalen Web Site hat. Die Häufigkeit und der Umfang der Erfassungsprozesse müssen frei konfigurierbar sein. Auf diese Weise lässt sich sicherstellen, dass das Analysieren der Web Site und das Generieren von Abbildern nur in Frequenzen stattfindet, die die Hard- und Software der abzubildenden Web Site verarbeiten kann.

### *Berücksichtigung der Interoperabilität einer Web Site*

In der Regel wird eine Web Site nicht alleinstehend betrieben, sondern kommuniziert mit weiteren IT-Systemen. Damit ein unabhängiges Archiv digitaler Abbilder entstehen kann, müssen sämtliche Abbilder von produktiven IT-Systemen der Organisation entkoppelt werden. Die Abbilder sollen losgelöst von der Systemumgebung der originalen Web Site bewahrt und bereitgestellt werden. Damit die Interoperabilität weiterhin für historische Auswertungen nachvollziehbar bleibt, können Systemschnittstellen als Bestandteil von

---

272 Vgl. Meidl, Oliver: Global Website: Webdesign im internationalen Umfeld, a. a. O., S. 52.

273 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 61 f.

274 Vgl. jQuery Foundation (Hrsg.): jQuery – write less, do more, a. a. O., 11.05.2018.

Kontext-Informationen bzw. Infrastruktur dokumentiert werden und auf diese Weise für historische Recherchen erhalten bleiben.

Nachdem die fachlichen Anforderungen im vorliegenden Kapitel formuliert und erläutert wurden, folgt die Definition der organisatorischen Anforderungen.

### 4.3 Organisatorische Anforderungsanalyse

#### 4.3.1 Systematisierung der organisatorischen Anforderungsanalyse

Neben den fachlichen Anforderungen werden im vorliegenden Kapitel organisatorische Anforderungen formuliert. Die organisatorischen Anforderungen formalisieren institutionelle Merkmale der organisatorischen Integrationsumgebung einer Web Site.<sup>275</sup>

Klassischerweise kann eine Organisation aus einer statischen und dynamischen Perspektive analysiert werden. Die statische Perspektive umfasst Strukturen, die dynamische Perspektive berücksichtigt Prozesse, die auf Basis struktureller Komponenten (Ressourcen wie Personal oder IT-Systeme) umgesetzt werden.<sup>276</sup>

Im vorliegenden Kapitel werden zunächst Anforderungen an die Struktur einer Organisation formuliert, anschließend folgen Anforderungen, die interne Prozesse der Organisation adressieren.

#### 4.3.2 Strukturorientierte Anforderungen

Die Struktur einer Organisation ist durch eine Gliederung und Zuweisung von Aufgaben, Verantwortlichkeiten und Kompetenzen auf Stellen innerhalb der Organisation charakterisiert.<sup>277</sup> Klassischerweise werden die Aufgaben, Verantwortlichkeiten und Kompetenzen auf eine Weise arrangiert, die bestmöglich zum Unternehmenserfolg bzw. zur Erreichung der Ziele einer Organisation beiträgt. Auch auf den erfolgreichen Einsatz eines IT-Systems trifft dies zu.<sup>278</sup>

---

275 Vgl. Balzert, Helmut: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, a. a. O., S. 5.

276 Vgl. Heinrich, Lutz J.; Heinzl, Armin; Roithmayr, Friedrich: Wirtschaftsinformatik-Lexikon: mit etwa 4000 Stichwörtern und 3700 Verweistichwörtern, a. a. O., S. 478 f.

277 Vgl. Bea, Franz Xaver; Göbel, Elisabeth: Organisation: Theorie und Gestaltung, 4., neu bearbeitete und erweiterte Auflage, Stuttgart: Lucius & Lucius 2010, S. 297.

278 Vgl. Bea, Franz Xaver; Göbel, Elisabeth: Organisation: Theorie und Gestaltung, a. a. O., S. 256 f.

IT-System-spezifische Aufgaben, Verantwortlichkeiten und Kompetenzen müssen geschaffen und an geeigneter Stelle organisatorisch verankert werden. In diesem Zusammenhang kann grundsätzlich zwischen der Integration in eine funktionale und in eine divisionale Organisationsstruktur unterschieden werden.<sup>279</sup>

Aufgaben, Kompetenzen und Verantwortlichkeiten zur Bereitstellung einer WSB sind bei einer funktionalen Organisation beispielsweise innerhalb einer dafür geeigneten (IT-)Abteilung anzuordnen. Bei einer divisionalen Organisation werden dagegen Funktionsabteilungen unter spezialisierten Objekten wie Geschäftsbereichen oder Produkten subsumiert. Unterhalb dieser Objekte entstehen typischerweise Divisionen, die sich erst anschließend in die o. g. Abteilungen oder vergleichbare Einheiten gliedern.<sup>280</sup> Im Regelfall ist bei einer divisionalen Organisationsform eine zentrale Institution vorhanden, die als IT-Service-Center agiert und die Integration von IT-Systemen übernimmt.

Da das WSBS als IT-System zur Dokumentation bestehender Informationstechnologie (Web Site) eingesetzt werden soll, wird in der vorliegenden Arbeit davon ausgegangen, dass grundlegende Strukturen zur Integration der Systemlösung vorhanden sind. Aufgaben, Kompetenzen und Verantwortlichkeiten sind ohne tiefgreifende Veränderungen in diesen Strukturen integrierbar. Herausforderungen entstehen vielmehr durch die Definition und Umsetzung von Aufgaben bzw. spezifischen Prozessen zum dauerhaften Einsatz dieser originären Systemlösung.

Aufbauend auf den strukturellen Anforderungen werden im folgenden Abschnitt Anforderungen an diese Prozesse beschrieben.

### 4.3.3 Prozessorientierte Anforderungen

In Ergänzung zur statischen Organisationsgestaltung lassen sich aus der Literatur organisatorische Anforderungen ableiten, die Prozesse zum Erstellen einer WSB beeinflussen bzw. bestimmen. Im Folgenden sollen diese Anforderungen beschrieben werden.<sup>281</sup>

In Anlehnung an klassische und digitale Archive, die sich zur Bewahrung digitaler Güter verpflichten, sind vier zentrale Eigenschaften feststellbar, die durch organisatorische

---

279 Vgl. Schreyögg, Georg; Geiger, Daniel: Organisation: Grundlagen moderner Organisationsgestaltung: mit Fallstudien, 6., vollständig überarbeitete und erweiterte Auflage, Wiesbaden: Springer Gabler 2016, S. 42 ff.

280 Vgl. Schreyögg, Georg; Geiger, Daniel: Organisation: Grundlagen moderner Organisationsgestaltung: mit Fallstudien, a. a. O., S. 44 ff.

281 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 50 ff.

Maßnahmen bzw. Prozesse erreicht werden müssen, um die Abbilder einer Web Site für die Nachwelt zu erhalten.<sup>282</sup>

- Rechtssicherheit
- Finanzierungssicherheit
- Kompetenzverfügbarkeit
- Qualitätssicherheit

Zum Sicherstellen bzw. Erreichen dieser Eigenschaften werden im Folgenden Anforderungen an die Organisation formuliert.

#### *Rechtssicherheit*

Einfluss auf die grundsätzliche Existenz und Beständigkeit einer WSB im Organisationsgefüge hat die Wahrung der Rechtssicherheit. Nur wenn das technische System innerhalb der akzeptierten Grenzen des wirkenden Rechtsraums agiert, kann der Betrieb langfristig gewährleistet und Sanktionen umgangen werden. Es sollen daher nicht nur qualifizierte Mitarbeiter vorhanden sein, sondern auch Prozesse implementiert werden, die sicherstellen, dass mit dem Einsatz des IT-Systems kein Verstoß gegen rechtliche Vorgaben erfolgt.<sup>283</sup>

Mitarbeiter müssen nicht nur reaktiv, sondern auch proaktiv auf Änderungen der Gesetzgebung reagieren und durch geeignete Maßnahmen notwendige Anpassungen an der Systemumgebung vornehmen. Es gilt, neben dem Datenschutz auch Anforderungen an die Langzeiterhaltung von digitalen Datenbeständen oder das Urheberrecht zu berücksichtigen. Sensible Inhalte auf der Web Site müssen mit Hilfe von Negativlisten exkludiert und im Idealfall proaktiv von der Erfassung ausgeschlossen werden. Inhalte, die von externen Anbietern bereitgestellt werden, sind in geeigneten Positivlisten zu erfassen. Die verlässliche und beständige Pflege umfassender Positiv- und Negativlisten ist entscheidend für die Regelkonformität des IT-Systems.

Nicht minder relevant ist zudem die Wahrung der Finanzierung. Es muss langfristig sichergestellt werden, dass das technische System sowie alle damit verbundenen Stellen in der Organisation finanziert und realisiert werden können.

---

282 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 50 ff.

283 Vgl. Brown, Adrian: Archiving websites: a practical guide for information management professionals, a. a. O., S. 146 ff. oder Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 51 f.

### *Finanzierungssicherheit*

Die Erhaltung der Abbilder einer Web Site ist mit finanziellen Aufwendungen verbunden. Die langfristige Bereitstellung und Pflege bedarf einer Budgetplanung, die eine finanzielle Versorgung innerhalb der Organisation sicherstellt. Durch das Schaffen und regelmäßige Prüfen notwendiger und verfügbarer Finanzierungsmöglichkeiten soll die Nachhaltigkeit des Systems auf organisatorischem Wege sichergestellt werden. Beispielhaft können Finanzierungen für den Betrieb, die Beschaffung von Hard- und Software sowie Personal und Schulungen angeführt werden. Im Idealfall ist in diesem Zusammenhang eine möglichst langfristige Prognose des benötigten finanziellen Budgets möglich.<sup>284</sup> Die Kosteneffizienz der technischen Bereitstellung ist darüber hinaus sorgfältig zu prüfen.<sup>285</sup>

### *Kompetenzverfügbarkeit*

In Ergänzung zur langfristigen finanziellen Sicherheit ist eine anhaltende Bereitstellung notwendiger Kompetenzen innerhalb der Organisation relevant. Es muss in geeignetem Umfang eine Planung, Steuerung und Kontrolle des Personals sichergestellt werden. Mitarbeiter mit geeigneten Qualifikationen müssen vorhanden sein, um das technische System zu nutzen und zu warten. Es gilt, die dauerhafte System-Verfügbarkeit durch fachliche Kompetenz zu wahren. In diesem Zusammenhang sind periodische oder episodische Prozesse zu berücksichtigen, die eine Weiter- bzw. Fortbildung umfassen.<sup>286</sup> Der anhaltende technologische Wandel geht mit kontinuierlichen Anpassungen von Funktionen und Dateiformaten im Web einher und macht Weiterbildungsmaßnahmen notwendig. Beispielhaft kann die unlängst erfolgte Implementierung des HTML-5-Standards oder die Abkehr von der Software „Flash“ angeführt werden. Darüber hinaus sollten Prozesse zur Krisenregelung definiert werden, die bei kritischem Personalmangel die Funktionalität des IT-Systems sicherstellen.<sup>287</sup>

---

284 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 52.

285 Vgl. Brown, Adrian: Archiving websites: a practical guide for information management professionals, a. a. O., S. 171.

286 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 52 f.

287 Vgl. z. B. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 55 f.

### *Qualitätssicherheit*

Vergleichbar mit der stetig zunehmenden Komplexität von IT-Systemen, setzt auch die Reproduktion der Lebensgeschichte einer Web Site Realisierungsmechanismen voraus, die über die Zeit an Komplexität gewinnen können.<sup>288</sup>

Eine Qualitätssicherung der technischen Komponenten zum Erzeugen einer WSB ist unabdingbar. Funktionalität, Zuverlässigkeit, Effizienz und Benutzbarkeit sollen durch periodische und ereignisgesteuerte Überprüfungen sichergestellt sein.<sup>289</sup>

Im Gegensatz zu syntaktischen Fehleranalysen, die sich beispielsweise beim Ausführen oder Kompilieren des geschriebenen Quellcodes äußern, zielt die vorliegende Anforderung vor allem auf das Überprüfen von Alterungsprozessen ab. Aus technischer Sicht sollen regelmäßige System-Tests sicherstellen, dass die Summe der Systemkomponenten in der Lage ist, vollständige Abbilder einer Web Site zu erzeugen. Nur so kann das Erlebarmachen der WSB langfristig erfolgen.<sup>290</sup>

In Ergänzung zur Formalisierung der organisatorischen Anforderungen, folgt im sich anschließenden Kapitel die Analyse technischer Anforderungen.

## 4.4 Technische Anforderungsanalyse

### 4.4.1 Systematisierung der technischen Anforderungsanalyse

Die zuvor skizzierten fachlichen und organisatorischen Anforderungen beschreiben funktionale Merkmale sowie Anpassungen organisatorischer Prozesse und Strukturen. Um eine vollständige Analyse der Anforderungen durchzuführen, müssen auch technische Eigenschaften berücksichtigt werden. Die technischen Eigenschaften sollen einen beständigen Betrieb der Systemlösung ermöglichen und werden im Regelfall aus standardisierten Qualitätsanforderungen abgeleitet. Qualitätsanforderungen sind meist allgemeingültig formuliert und umfassen somit Kriterien, die für sämtliche Arten von IT-Systemen notwendig und anwendbar sind, auch für das WSBS.

---

288 Vgl. Hoffmann, Dirk W.: Software-Qualität, Berlin: Springer 2008, S. 12 f.

289 Vgl. Hoffmann, Dirk W.: Software-Qualität, a. a. O., S. 6 f.

290 Vgl. auch Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 49 f.

Abgeleitet aus der ISO-Norm 9126 gilt es die in Abbildung 15 dargestellten Qualitätsanforderungen für ein WSBS zu berücksichtigen.<sup>291</sup>

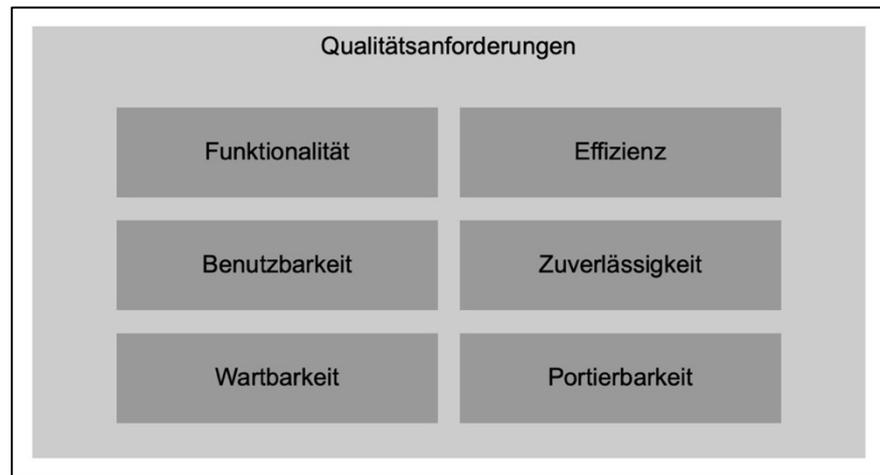


Abb. 15: Qualitätsanforderungen an ein klassisches AWS

Eine Spezifikation der dargestellten Anforderungen erfolgt in den nachstehenden Abschnitten des vorliegenden Kapitels.<sup>292</sup>

Da standardisierte Qualitätsanforderungen einen generischen Charakter aufweisen, adressieren sie keine spezifischen Herausforderungen der digitalen Erhaltung. Im Rahmen der technischen Anforderungsanalyse ist es daher notwendig, ergänzende Anforderungen zu definieren, die aus technischer Perspektive den Erhalt digitaler Artefakte adressieren. Der zweite Abschnitt der technischen Anforderungsanalyse widmet sich deshalb der Beschreibung von Anforderungen, die der Beschädigung von Hardware (*Hardware-Korruption*), der Beschädigung von Dateien (*File-Korruption*) und der Überalterung von Dateiformaten (*Format-Obsoleszenz*) entgegenwirken sollen.<sup>293</sup>

#### 4.4.2 Standardisierte Qualitätsanforderungen

Qualitätsanforderungen beschreiben technische Aspekte, die in der Regel ein Set an Funktionen oder den kompletten Funktionsumfang eines IT-Systems betreffen. Ziel der Qualitätsanforderungen ist das Sicherstellen der Betriebsfähigkeit einer Systemlösung.

291 Vgl. Bergsmann, Johannes; Unterauer, Markus: Requirements Engineering für die agile Softwareentwicklung: Methoden, Techniken und Strategien, a. a. O., S. 107 f.

292 Vgl. z. B. auch Rupp, Chris: Requirements-Engineering und -Management: professionelle, iterative Anforderungsanalyse für die Praxis, 5., aktualisierte und erw. Aufl, München: Hanser 2009, S. 260 ff.

293 Vgl. Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften, DARIAH-DE (Hrsg.): Handbuch Digital Humanities: Anwendungen, Forschungsdaten und Projekte / DARIAH-DE, a. a. O., S. 63.

Sie haben einen unmittelbaren Einfluss auf die IT-Architektur und sind in standardisierten Normen festgehalten.<sup>294</sup>

Standardisierte Qualitätsanforderungen auf Basis der ISO-Norm 9126 umfassen: Funktionalität, Effizienz, Zuverlässigkeit, Benutzbarkeit, Wartbarkeit und Portierbarkeit.<sup>295</sup>

### *Funktionalität*

Die Forderung nach Funktionalität macht explizit, dass alle geforderten Funktionen technisch und vollständig realisiert sein müssen. Das WSBS soll alle Funktionen fehlerfrei bereitstellen, um eine WSB dauerhaft zu erzeugen. Sämtliche Systembausteine, die zur Realisierung eingesetzt werden, müssen funktionieren. Der funktionale Umfang soll die fachlichen Anforderungen vollständig abbilden und zur Verfügung stehen.<sup>296</sup>

### *Effizienz*

Die Forderung nach Effizienz wird im Wesentlichen durch ein angemessenes Verhältnis zwischen Leistungsniveau und notwendigen bzw. eingesetzten Ressourcen charakterisiert. Der Betrieb soll auf Basis von Hardware-Komponenten durchgeführt werden, die optimierte Antwort- bzw. Ladezeiten, einen angemessenen Betriebsmitteleinsatz und skalierbare Ressourcen ermöglichen.<sup>297</sup>

### *Zuverlässigkeit*

Neben einem effizienten Einsatz der verfügbaren Ressourcen soll ein IT-System zuverlässig genutzt werden können. Es soll möglichst keine Systemausfälle aufweisen und eine dauerhafte Verfügbarkeit der Funktionen berücksichtigen. Eine Skalierbarkeit der Systemressourcen ist zudem notwendig, da zukünftige Entwicklungen, Herausforderungen und Anforderungen an den Funktionsumfang mit zunehmenden Systemanforderungen verbunden sein können.

---

294 Vgl. Balzert, Helmut: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, a. a. O., S. 463 f.

295 Vgl. Bergsmann, Johannes; Unterauer, Markus: Requirements Engineering für die agile Softwareentwicklung: Methoden, Techniken und Strategien, a. a. O., S. 107 f.

296 Vgl. Balzert, Helmut: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, a. a. O., S. 468 oder Bergsmann, Johannes; Unterauer, Markus: Requirements Engineering für die agile Softwareentwicklung: Methoden, Techniken und Strategien, a. a. O., S. 110.

297 Vgl. Bergsmann, Johannes; Unterauer, Markus: Requirements Engineering für die agile Softwareentwicklung: Methoden, Techniken und Strategien, a. a. O., S. 109.

Eine geringe Fehlertoleranz soll darüber hinaus mit einer schnellst möglichen Wiederherstellbarkeit des Leistungsniveaus bei Ausfällen einzelner oder aller Funktionen einhergehen. Das Auftreten von Fehlerzuständen soll nicht zum Absturz des Gesamtsystems führen. Ein geeignetes Fehlermanagement ist daher zu implementieren.<sup>298</sup>

### *Benutzbarkeit*

Aus Sicht der Anwender markiert die Benutzerbarkeit einen zentralen Aspekt aus technischer Perspektive. Für den Erfolg und die langfristige Etablierung innerhalb der Organisation, ist die Verständlichkeit und Erlernbarkeit der zur Verfügung gestellten Funktionen unerlässlich. Benutzer müssen in der Lage sein, die notwendigen Steuerungselemente und Bedienoberflächen zu verstehen und zu verwenden. In diesem Zusammenhang sollte eine attraktive Ausgestaltung der Benutzeroberflächen nicht vernachlässigt werden.<sup>299</sup> Im Gegensatz zu klassischen Archiven soll die Systemlösung nicht nur die Nachnutzbarkeit alternder Dateiformate in den Fokus stellen, sondern vor allem das Erleben der Vergangenheit auf performanten Web-Oberflächen.<sup>300</sup>

### *Wartbarkeit*

Das Qualitätskriterium der Wartbarkeit beschreibt die Fähigkeit eines IT-Systems, auf notwendige, gewollte oder ungewollte Veränderungen zu reagieren. Um eine möglichst effiziente Wartung zu ermöglichen, soll eine modulare Entwicklung berücksichtigt werden. Das Erfassen, Bewahren und Bereitstellen einer WSBS sollen durch abgeschlossene Systembausteine realisiert werden. Die Veränderung eines Systembausteins soll möglichst keine Auswirkungen auf andere Systembausteine haben.<sup>301</sup>

### *Portierbarkeit*

Als abschließende Anforderung ist die Portierbarkeit zu erwähnen. Das IT-System soll möglichst flexibel auf neue Betriebsumgebungen angepasst werden können. Sowohl Software als auch Hardware zum Betrieb unterliegen einem kontinuierlichen Wandel. Es ist daher wesentlich, dass technische Änderungen ohne Restriktionen durchgeführt wer-

---

298 Vgl. Bergsmann, Johannes; Unterauer, Markus: Requirements Engineering für die agile Softwareentwicklung: Methoden, Techniken und Strategien, a. a. O., S. 112.

299 Vgl. Balzert, Helmut: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, a. a. O., S. 469.

300 Vgl. Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften, DARIAH-DE (Hrsg.): Handbuch Digital Humanities: Anwendungen, Forschungsdaten und Projekte / DARIAH-DE, a. a. O., S. 67.

301 Vgl. Bergsmann, Johannes; Unterauer, Markus: Requirements Engineering für die agile Softwareentwicklung: Methoden, Techniken und Strategien, a. a. O., S. 150.

den können. Es soll eine Nutzung auf wechselnden Betriebssystemen und eine restriktionsfreie Koexistenz mit sonstiger organisationaler Software und Hardware erreicht werden.<sup>302</sup>

Die oben genannten Qualitätsanforderungen geben Auskunft über die notwendigen Standards der Hard- und Software. Das nachfolgende Kapitel erweiterte diese Anforderungen um ein spezialisiertes Set an technischen Anforderungen, die sich unmittelbar der Bestandserhaltung einer WSBS widmen.

#### 4.4.3 Anforderungen an den Erhalt digitaler Abbilder

Nachdem standardisierte Qualitätsanforderungen formuliert wurden, folgt die Darlegung spezieller technischer Anforderungen, die Aspekte für den langfristigen und physischen Erhalt digitaler Abbilder einer Web Site berücksichtigen.

In diesem Zusammenhang sind Systemeigenschaften relevant, die auf eine Vermeidung von Hardware-Korruption, File-Korruption und Format-Obsoleszenz abzielen.<sup>303</sup>

Um diese Risiken zu reduzieren, sind aus technischer Sicht folgende Systemeigenschaften zu berücksichtigen:<sup>304</sup>

- Redundanz und Verteilung
- Diversifikation und Standardisierung
- Medienmigration
- Emulation

##### *Redundanz und Verteilung*

Zentraler Bestandteil einer ausfallsicheren Systemumgebung ist eine redundante Sicherung der Abbilder. Die redundante Sicherung der Abbilder beugt dem Verlust von Daten durch Ausfall oder Beschädigung einzelner Datenträger vor.

Um auf negative Umwelteinflüsse zu reagieren, sollte nicht nur eine Verteilung bzw. Vielfältigung der Daten auf mehreren Datenträgern berücksichtigt werden, sondern auch eine geographische bzw. räumliche Aufgliederung auf unterschiedliche Standorte. Der

---

302 Vgl. Balzert, Helmut: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, a. a. O., S. 470.

303 Vgl. Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften, DARIAH-DE (Hrsg.): Handbuch Digital Humanities: Anwendungen, Forschungsdaten und Projekte / DARIAH-DE, a. a. O., S. 63.

304 Vgl. Ullrich, Dagmar: Bitstream Preservation, a. a. O., S. 165 f.

Gefahr eines Datenverlustes durch menschliches Versagen kann zudem durch eine Aufteilung der Datenspeicher auf mehrere, unabhängige Organisationseinheiten (z. B. Drittanbieter) entgegengewirkt werden.<sup>305</sup>

#### *Diversifikation und Standardisierung*

Abbilder einer Web Site sollen auf Datenträger verteilt werden, um eine möglichst ausfallsichere Bewahrung durch redundante Speicherung zu erreichen. Der Einsatz unterschiedlicher bzw. abweichender Speichertechniken soll in diesem Zusammenhang berücksichtigt werden.<sup>306</sup>

Das WSBS soll eine Diversifizierung der eingesetzten Speichertechniken ermöglichen, um von den Vorteilen unterschiedlicher Datenträger zu profitieren. Das Risiko der Unbrauchbarkeit von Abbildern durch Alterung kann so gemindert werden. Darüber hinaus sollen Datenträger und Speichertechniken eine Standardisierung und internationale Verbreitung aufweisen, um eine langfristige Verfügbarkeit und Support zu gewährleisten.<sup>307</sup>

#### *Medienmigration*

In Ergänzung zu der Diversifizierung von Speichertechniken soll an dieser Stelle die Forderung nach einer beständigen Medienmigration erläutert werden. Grundsätzlich kann bei der Medienmigration zwischen vier verschiedenen Verfahren unterschieden werden: Einem Erneuerungs-, Replikations-, Restrukturierungs- und Transformationsverfahren. Bei einem Erneuerungsverfahren wird im Regelfall ein älterer Datenträger durch einen neuen aber gleichartigen Datenträger ausgetauscht. Der Bestand an Inhaltsdaten wird dabei ohne weitere Anpassungen auf das neue Trägermedium kopiert. Das Replikationsverfahren setzt dagegen auf das Austauschen von Datenträgern, die nicht zwingend gleichartig sind. Es werden unter Umständen vollständig neue Trägermedien mit abweichender Speichertechnik integriert. Dabei sind durchaus aufwendige Anpassungen der Speicherinfrastruktur zu berücksichtigen, da der neue Datenträger nicht unmittelbar als Substitut verwendet und alle Inhalte restriktionsfrei auf das neue Medium kopiert werden können. Im Restrukturierungsverfahren findet dagegen ausschließlich eine Modifikation der Struktur der Daten statt, die Inhalte selbst werden nicht angepasst. Beispielhaft kann an dieser Stelle die Umwandlung von ZIP-Dateien in (Win)RAR-Dateien angeführt werden. Das Transformationsverfahren setzt tiefgreifende Restrukturierungsmaßnahmen voraus. Bei

---

305 Vgl. Ullrich, Dagmar: Bitstream Preservation, a. a. O., S. 166. oder Rathje, Ulf: Technisches Konzept für die Datenarchivierung im Bundesarchiv, in: Der Archivar, 2/2002, S. 118 ff.

306 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 20 f. oder Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 46 f.

307 Vgl. Ullrich, Dagmar: Bitstream Preservation, a. a. O., S. 116.

diesem Migrationsverfahren werden die digitalen Inhalte selbst formatiert.<sup>308</sup> Umwandlung eines Bildes im Dateiformat JPEG in das Dateiformat PNG oder die Transformation einer HTML-Datei in eine XML-Datei können hier beispielhaft genannt werden.<sup>309</sup>

### *Emulation*

Neben einem Migrationsverfahren sollte eine Emulationsumgebung vorhanden sein. Als Alternative zum Migrationsverfahren forciert eine Emulationsumgebung nicht das Verändern der tatsächlichen Inhalte, sondern stellt eine IT-Umgebung zur Verfügung, die veraltete Inhalte in historischen Umgebungen nutzbar macht. Für alternde Software soll langfristig eine Emulationsumgebung vorgehalten werden, die sicherstellt, dass veraltete Dateien auch in Zukunft abgebildet und genutzt werden können.<sup>310</sup>

## 4.5 Zusammenfassung und Erläuterung des Fortgangs der Arbeit

Die Anforderungsanalyse gliedert sich in fachliche, organisatorische und technische Anforderungen. Die fachlichen Anforderungen beschreiben zunächst den funktionalen Umfang zur systembasierten Realisierung einer WSB.

Vergleichbar mit Systemanalysen, die Anforderungen an digitale Archive auf Basis ihrer Risikofaktoren formalisieren<sup>311</sup>, wurden im vorliegenden Kapitel fachliche Anforderungen auf Basis der Eigenschaften einer Web Site analysiert und definiert.<sup>312</sup>

Die Eigenschaften einer Web Site sind nach inhaltlichen, gestalterischen und betriebspezifischen Aspekten gegliedert, um eine umfassende Beschreibung des Konstruktionsziels der Arbeit zu gewährleisten.<sup>313</sup>

---

308 Vgl. Funk, Stefan E.: Migration, in: nestor Handbuch: eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 2.3, 2010, Hrsg.: Neuroth, Heike; Obwald, Achim; Scheffel, Regine; Strathmann, Stefan; Jehn, Mathias, 2., aktualisierte Auflage, Glückstadt: Verlag Werner Hülsbusch, Fachverlag für Medientechnik und -wirtschaft 2016, S. 157.

309 Vgl. z. B. Ullrich, Dagmar: Bitstream Preservation, a. a. O., S. 117 f. oder Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 52 ff.

310 Vgl. Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 54 ff.

311 Vgl. Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften, DARIAH-DE (Hrsg.): Handbuch Digital Humanities: Anwendungen, Forschungsdaten und Projekte / DARIAH-DE, a. a. O., S. 63.

312 Vgl. Rosenthal, David; Robertson, Thomas; Lipkis, Tom; Reich, Vicky; Morabito, Seth: Requirements for Digital Preservation Systems – A Bottom-Up Approach, a. a. O., 12.08.2017.

313 Vgl. z. B. Zschau, O.; Traub, D.; Zahradka, R.: Web Content Management. Websites professionell planen und betreiben. 2., überarb. und erw. Aufl, a. a. O., S. 169.

Konsequenterweise wird eine organisatorische bzw. technische Anforderungsanalyse angefügt. Organisatorische und technischen Anforderungen adressieren nicht-funktionale Eigenschaften, die zum Betreiben eines IT-Systems notwendig sind.<sup>314</sup> Die organisatorischen und technischen Anforderungen beschreiben die Qualität betriebsnotwendiger Komponenten und erläutern Prozesse bzw. Maßnahmen, die innerhalb einer Organisation implementiert werden müssen, um einen dauerhaften Betrieb zu ermöglichen.<sup>315</sup>

Aufbauend auf der Analyse der Anforderungen widmet sich das folgende Kapitel 5 der konzeptionellen Gestaltung eines WSBS. Im Rahmen eines Systementwurfs werden eine Systemarchitektur und einzelne Systembausteine skizziert, die für die Umsetzung eines WSBS relevant sind. Die Summe der Systembausteine beschreibt die geforderte Systemlösung – das Web-Site-Biographie-System (kurz: WSBS).

---

314 Vgl. Balzert, Helmut: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, a. a. O., S. 489.

315 Vgl. Balzert, Helmut: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, a. a. O., S. 456 ff.

## 5 Systementwurf eines WSBS

### 5.1 Systematisierung des Systementwurfs

Der Systementwurf beschreibt ein IT-System aus technisch-konzeptioneller Perspektive. Durch eine Zerlegung in abgeschlossene Systemmodule entsteht ein strukturierter Bauplan. Die Summe der Systemmodule beschreibt eine ganzheitliche Systemlösung.<sup>316</sup>

Für die Entwicklung eines Systementwurfs lassen sich in der Fachliteratur vier etablierte Entwurfsverfahren feststellen, die im Ergebnis ein IT-System beschreiben: Das Top-Down-, Bottom-Up-, Outside-In- und Inside-Out-Verfahren.<sup>317</sup>

Abbildung 16 stellt die Verfahren zum Realisieren eines Systementwurfs schematisch und im Überblick dar.

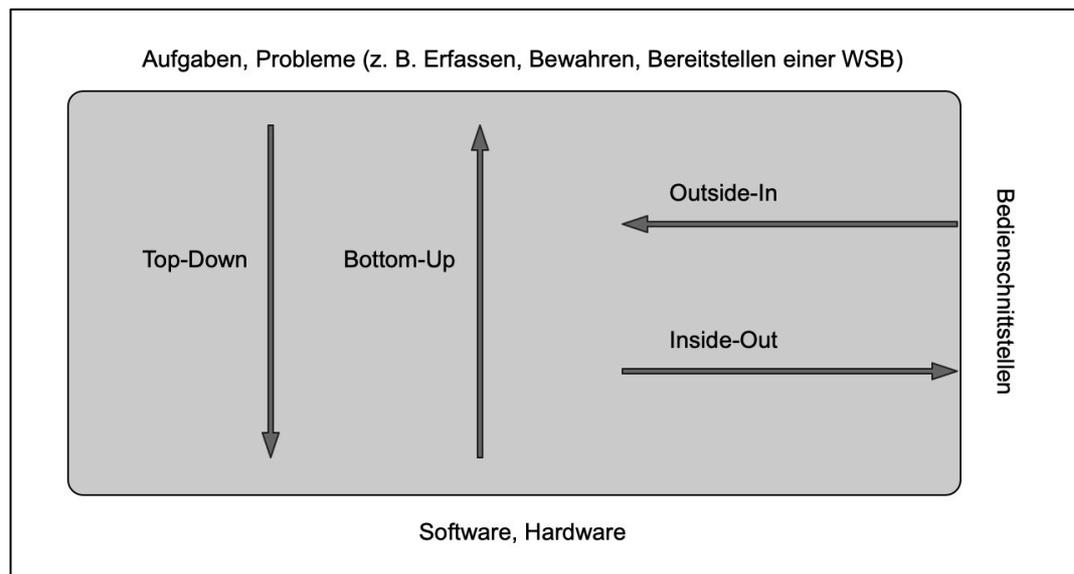


Abb. 16: Vorgehensweisen des Systementwurfs

316 Vgl. Ludewig, Jochen; Lichter, Horst: Software Engineering: Grundlagen, Menschen, Prozesse, Techniken, a. a. O., S. 400.

317 Vgl. Balzert, Helmut: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, a. a. O., S. 54 ff. oder Ludewig, Jochen; Lichter, Horst: Software Engineering: Grundlagen, Menschen, Prozesse, Techniken, a. a. O., S. 402 und Vogel, Oliver (Hrsg.): Software-Architektur: Grundlagen - Konzepte - Praxis, 2. Aufl, Heidelberg: Spektrum Akad. Verl 2009, S. 38.

Das Top-Down-Verfahren setzt bei generischen Aufgaben bzw. Problemen an und verfeinert diese, bis eine möglichst effiziente und effektive Implementierung auf Software- und Hardware-Ebene möglich ist. Das Bottom-Up-Verfahren geht dagegen von einer definierten Hardware- und Software-Ausstattung aus und leitet auf Grundlage dieser Ausstattung Probleme oder Aufgaben ab, die mit der vorhandenen Technik gelöst werden. Sind Bedienoberflächen durch den Kunden bereits definiert, findet das Outside-In-Verfahren eine Anwendung. Wird ausgehend von Aufgaben, Problemen sowie Hard- und Software-Kombinationen ein Bedienoberflächenentwurf abgeleitet, erfolgt ein Systementwurf nach dem Inside-Out-Verfahren.<sup>318</sup>

Für den vorliegenden Systementwurf sind die Aufgaben bzw. Probleme durch Anforderungen konkretisiert. Es wird daher zunächst das Top-Down-Verfahren zum Beschreiben der Software- und Hardware-Elemente gewählt. Im Rahmen einer Systemarchitektur werden Systembausteine festgelegt, die zur Realisierung der Anforderungen verwendet werden. Aufbauend auf der Systemarchitektur folgt eine konzeptionelle Verfeinerung.<sup>319</sup>

Zur Verfeinerung des Systementwurfs werden Spezifikationen der Systembausteine vorgenommen. Konkret wird auf Datenstrukturen, Datenbanksysteme, Anwendungen und Funktionen zum Erfassen, Bewahren und Bereitstellen einer WSB eingegangen. Die Funktionen resultieren aus den Anforderungen und werden in Modulen konsolidiert. Die Summe der Module beschreibt den vollständigen Funktionsumfang.<sup>320</sup>

Ergänzend zum technischen und funktionalen Innenleben werden gemäß des Inside-Out-Verfahrens Bedienoberflächen abgeleitet.<sup>321</sup> Die Bedienoberflächen werden als Skizzen entworfen und visualisieren Strukturen zum Verwalten und Erleben der Lebensgeschichte einer Web Site.

---

318 Vgl. Ludewig, Jochen; Lichter, Horst: Software Engineering: Grundlagen, Menschen, Prozesse, Techniken, a. a. O., S. 402 f.

319 Vgl. die Definition von Modulen in Ludewig, Jochen; Lichter, Horst: Software Engineering: Grundlagen, Menschen, Prozesse, Techniken, a. a. O., S. 405.

320 Vgl. Balzert, Helmut: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, a. a. O., S. 40 oder auch Bergsmann, Johannes; Unterauer, Markus: Requirements Engineering für die agile Softwareentwicklung: Methoden, Techniken und Strategien, a. a. O., S. 105 f.

321 Vgl. Ludewig, Jochen; Lichter, Horst: Software Engineering: Grundlagen, Menschen, Prozesse, Techniken, a. a. O., S. 402 f.

## 5.2 Systemarchitektur: Systembausteine eines WSBS

### 5.2.1 Systematisierung der Systemarchitektur

Die Systemarchitektur ist eine technische Blaupause für den Prototypen. Sie umfasst den Aufbau von Hardware- und Software-Komponenten, die zur Realisierung geforderter Funktionen notwendig sind.<sup>322</sup> Für eine WSB werden sämtliche Funktionen mit dem WSBS bereitgestellt. Die Systemarchitektur des WSBS wird in den nachfolgenden Kapiteln skizziert.

Eine Systemarchitektur gibt Aufschluss über Beziehungen und Schnittstellen, technische Verantwortlichkeiten sowie Aufgabenbereiche, Gruppierungen, Schichten und die Verteilung einzelner Systembausteine.<sup>323</sup>

Für die stringente Beschreibung der Systembausteine lassen sich unterschiedlich etablierte und in ihrer Ausprägung variierende Modelle feststellen, die je nach Einsatzbereich des IT-Systems geeignet sein können.<sup>324</sup>

Um ein umfassendes Modell zu entwickeln, wird in den folgenden Kapiteln eine Strukturierung in drei Ebenen vorgenommen:<sup>325</sup> Die drei Ebenen lassen sich aus der logischen Gliederung eines IT-Systems in eine Datenhaltungs-, Steuerungs- und Präsentationsschicht ableiten.<sup>326</sup>

Zur Datenhaltungsschicht werden Systembausteine beschrieben, die sich mit dem persistenten Speichern und Bereitstellen einer WSB befassen. Es wird auf Dateiformate und Datenbanksysteme eingegangen. Klassischerweise werden in diesem Kontext relationale oder nicht-relationale Datenbanksysteme analysiert und Datenmodelle zur Implementierung beschrieben.

---

322 Vgl. Dern, Gernot: Management von IT-Architekturen: Leitlinien für die Ausrichtung, Planung und Gestaltung von Informationssystemen, 2., verb.erw. Aufl, Wiesbaden: Vieweg 2006, S. 28 f.

323 Vgl. Vogel, Oliver; Arnold, Ingo; Chughtai, Arif; Ihler, Edmund; Kehrer, Timo; Mehlig, Uwe; Zdun, Uwe: Software-Architektur: Grundlagen - Konzepte - Praxis, 2. Aufl, Heidelberg: Spektrum Akad. Verl 2009, S. 8 f.

324 Vgl. Vogel, Oliver; Arnold, Ingo; Chughtai, Arif; Ihler, Edmund; Kehrer, Timo; Mehlig, Uwe; Zdun, Uwe: Software-Architektur: Grundlagen - Konzepte - Praxis, a. a. O., S. 216 ff.

325 Vgl. Ludewig, Jochen; Lichter, Horst: Software Engineering: Grundlagen, Menschen, Prozesse, Techniken, a. a. O., S. 400 oder auch Vogel, Oliver (Hrsg.): Software-Architektur: Grundlagen - Konzepte - Praxis, a. a. O., S. 217 ff.

326 Vgl. Heide Balzert: UML 2 in 5 Tagen: der schnelle Einstieg in die Objektorientierung, 2. Aufl., 1. Nachdruck, Herdecke: W3L-Verl 2010, S. 123 f.

Zur Beschreibung der Steuerungsschicht werden Anwendungen erläutert, die auf Basis der Datenschicht relevante Funktionsmengen nutzbar machen. Anwendungen werden definiert und Einsatzgebiete bzw. Verantwortlichkeiten dieser Anwendungen als Bestandteil des WSBS erläutert.

Typische Systembausteine zur Steuerung des IT-Systems sind – neben Web Servern und Firewalls – sämtliche Programme, die das Erfassen, Bewahren und Bereitstellen der WSB ermöglichen.<sup>327</sup>

Grundlegend für die Bereitstellung einer WSB ist die Präsentationsschicht. Im Systementwurf werden Skizzen von Bedienoberflächen erzeugt, die im Web Browser eines Nutzers umgesetzt werden und eine WSB erlebbar machen. Bei den Bedienoberflächen spielt das Client-System eine entscheidende Rolle. Inhalte und Gestaltungselemente einer WSB werden über das Internet transportiert und im Web Browser visualisiert.<sup>328</sup>

Wurden in den Anfängen des WWW Bedienoberflächen hauptsächlich mit HTML-Code erzeugt, gibt es heute ergänzende Techniken, die zum Realisieren von interaktiven Bedienoberflächen eingesetzt werden. Auf Grundlage von JavaScript-Frameworks lassen sich beispielsweise Bedienoberflächen entwickeln, deren Funktionsumfang vergleichbar mit klassischen Desktop-Anwendungen ist.<sup>329</sup>

Aufbauend auf der Systematisierung der Systemarchitektur werden im folgenden Kapitel die einzelnen Systembausteine dargestellt. Anschließend folgt eine detaillierte Beschreibung der Datenhaltungs-, Steuerungs- und Präsentationsschicht.

---

327 Vgl. Vogel, Oliver; Arnold, Ingo; Chughtai, Arif; Ihler, Edmund; Kehrer, Timo; Mehlig, Uwe; Zdun, Uwe: Software-Architektur: Grundlagen - Konzepte - Praxis, a. a. O., S. 224 f.

328 Vgl. Vogel, Oliver; Arnold, Ingo; Chughtai, Arif; Ihler, Edmund; Kehrer, Timo; Mehlig, Uwe; Zdun, Uwe: Software-Architektur: Grundlagen - Konzepte - Praxis, a. a. O., S. 223 ff.

329 Vgl. Steyer, Manfred; Softic, Vildan: AngularJS: moderne Webanwendungen und Single Page Applications mit JavaScript, Beijing Cambridge Farnham Köln Sebastopol Tokyo: O'Reilly 2015, S. 3 ff. oder Vogel, Oliver; Arnold, Ingo; Chughtai, Arif; Ihler, Edmund; Kehrer, Timo; Mehlig, Uwe; Zdun, Uwe: Software-Architektur: Grundlagen - Konzepte - Praxis, a. a. O., S. 227 f.



Im Datenbank- und Dateisystem erfolgt das persistente Speichern und Erhalten der WSB. Das Datenbanksystem ist klassischerweise mit Verzeichnissen in einem Dateisystem kombiniert, die große oder proprietäre Dateien außerhalb der Datenbanken bewahren.<sup>330</sup>

Das Befüllen der Datenbank selbst erfolgt durch die Crawler-Anwendung. Aufgabe der Crawler-Anwendung ist das Erfassen einer WSB. Abbilder werden durch die Crawler-Anwendung analysiert und bei Bedarf generiert. Das Bewahren und Bereitstellen ist anschließend zentraler Aufgabenbestandteil der Curator-Anwendung.<sup>331</sup> Mit Hilfe der Curator-Anwendungen können Administratoren Abbilder suchen, aufbereiten, manipulieren, verwalten und erhalten. Die Curator-Anwendung stellt alle Funktionen bereit, die ein Management digitaler Abbilder sowie das Erhalten und Zugänglichmachen dieser digitalen Artefakte ermöglichen.<sup>332</sup> Das Zusammenstellen und Präsentieren der Abbilder erfolgt im Web Browser. Die Abbilder einer Web Site werden auf Basis geeigneter Bedienoberflächen erlebbar gemacht.

Eine Bedienoberfläche für Nutzer ermöglicht das Erleben der WSB und stellt interaktive Präsentations- und Navigationsumgebungen bereit. Abbilder einer Web Site werden in standardisierten Formaten (z. B. JSON, XML) ausgeliefert und im Web Browser interpretiert.<sup>333</sup> Anpassungen auf unterschiedliche Endgeräte lassen sich durch einen standardisierten Output flexibel und effizient realisieren.<sup>334</sup>

Eine Bedienoberfläche für Administratoren stellt Funktionen zur Systemverwaltung bereit. Auf Basis dieser Funktionen können Administratoren die Crawler-Anwendung konfigurieren und Abbilder einer Web Site mit Hilfe der Curator-Anwendung verwalten.

Verfügbarkeitsdienste werden zunächst keiner spezifischen Schicht zugeordnet, da diese Dienste ergänzend zu den restlichen Systembausteinen fungieren. Sie stellen schichtenübergreifend die Verfügbarkeit der Systemlösung sicher.

Aufbauend auf der überblicksartigen Beschreibung der Systembausteine folgt eine detaillierte Auseinandersetzung mit technischen Bestandteilen der Datenhaltungs-, Steuerungs- und Präsentationsschicht.

---

330 Vgl. Schicker, Edwin: Datenbanken und SQL: eine praxisorientierte Einführung mit Anwendungen in Oracle, SQL Server und MySQL, 5., aktualisierte und erweiterte Auflage, Wiesbaden: Springer Vieweg 2017, S. 225 ff.

331 Vgl. zum Begriff „Curator-Anwendung“ den Begriff „Curation“ in Harvey, D. R.; Gillian, Oliver: Digital curation, a. a. O., S. 5 ff.

332 Vgl. Harvey, D. R.; Gillian, Oliver: Digital curation, a. a. O., S. 7.

333 Vgl. Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 58 f.

334 Vgl. z. B. Patni, Sanjay: Pro RESTful APIs: Design, Build and Integrate with REST, JSON, XML and JAX-RS, a. a. O., S. 11 ff.

### 5.2.3 Datenhaltungsschicht

Für den Systementwurf der Datenhaltungsschicht wird der Einsatz des Datenbank- und Dateisystems erläutert (vgl. Abb. 18 zur Einordnung der Systembausteine).

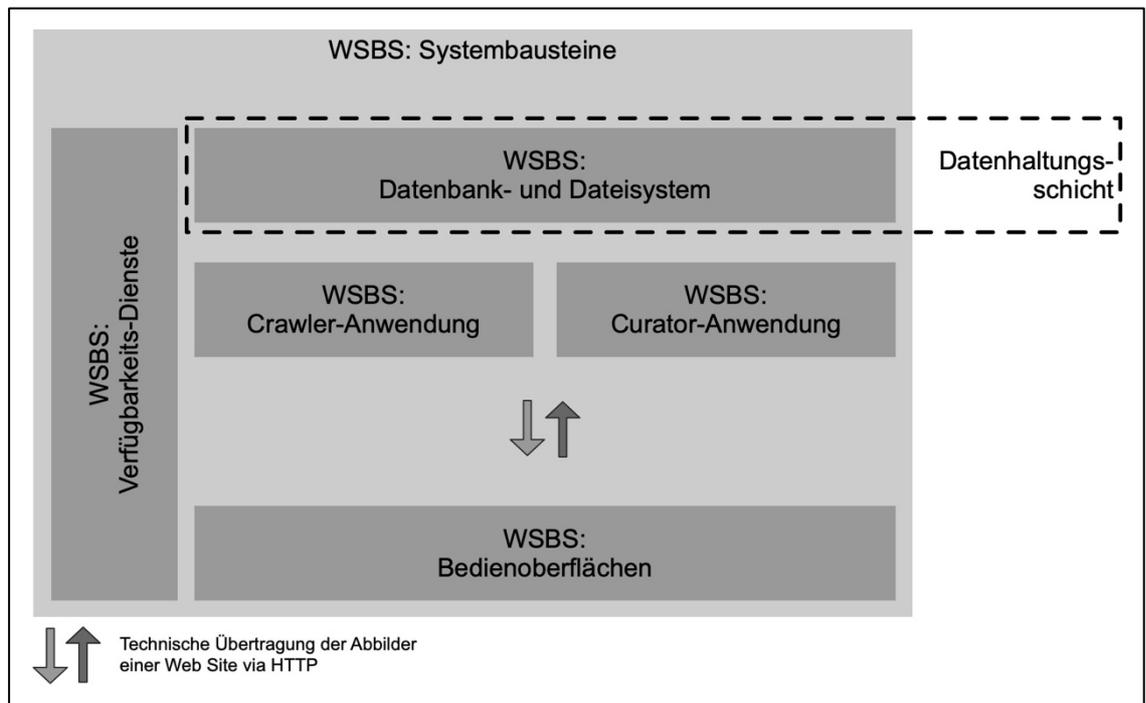


Abb. 18: Systembausteine der Datenhaltungsschicht

Zur dauerhaften Speicherung einer WSB müssen Abbilder einer Web Site in einem Datenbank- bzw. Dateisystem abgelegt werden.<sup>335</sup> Um den langfristigen Erhalt zu gewährleisten sind geeignete Dateiformate zu wählen, die den binären Zeichenstrom für die Nachwelt aufbewahren und interpretierbar machen.<sup>336</sup>

Bei der Auswahl geeigneter Dateiformate kann ein Zielkonflikt zwischen der Aufbewahrung in Container-Dateien wie ZIP- oder WARC-Dateien und der Sicherung in standardisierten Formaten wie XML, JSON oder HTML attestiert werden.<sup>337</sup> Im Folgenden soll auf die Vor- und Nachteile der einzelnen Formate eingegangen werden, um anschließend eine begründete Auswahl zu treffen.

335 Vgl. Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 121 ff.

336 Vgl. Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 58 ff.

337 Vgl. Masanès, Julien: Web Archiving: Issues and Methods, a. a. O., S. 33 f.

Das Verwenden von Container-Dateien ist üblicherweise mit vergleichsweise langen Verarbeitungszeiten verbunden, da mehrere Dateien in einer Container-Datei zusammengefasst sind und extrahiert werden müssen. Nur so kann eine Anwendung diese interpretieren. Gerade für Initiativen, die das WWW als Ganzes sichern, bietet sich die Bewahrung in Container-Dateien trotz der längeren Verarbeitungszeiten an. Große Datenmengen werden in kompakter Form gespeichert und damit Systemressourcen im Sinne von Speicherplatz geschont. Bei Initiativen, die sich mit einer spezifischen Web Site auseinandersetzen, führt das Sichern nicht-gepackter Einzeldateien zu einer effizienteren Verwendung. Im Gegensatz zu einer Container-Datei lassen sich einzelne Dateien schneller verarbeiten und in Datenbanken abbilden, da kein Prozess des Extrahierens oder Verpackens vorausgesetzt wird.

Mit dem WSBS soll eine Systemlösung geschaffen werden, die zum Erlebbarmachen einer WSB eingesetzt wird. Nutzerfreundlichkeit und Performance stehen dabei im Fokus. Im Gegensatz zu Archivsystemen, die Abbilder in enormen sowie redundanten Mengen speichern, soll die WSB Ladezeiten aufweisen, die mit der originalen Web Site vergleichbar sind.<sup>338</sup> Das technische Fundament zur Bewahrung bildet daher im Idealfall ein leistungsstarkes Datenbanksystem zum Abbilden von Einzeldateien.<sup>339</sup> Die Anzahl und Größe der zur Verfügung stehenden Datenbanken sollte skalierbar sein, um auf einen wachsenden Speicherbedarf zu reagieren.<sup>340</sup>

Neben Datenbanken ist ein Dateisystem zu implementieren.<sup>341</sup> Aus Performance-Gründen sollten nicht alle Dateien in Datenbanken gespeichert werden. Bei einer großen Datei oder einem proprietären Dateiformat enthalten Datenbanken Referenzpfade, die auf ein Verzeichnis verweisen, in dem die Datei abgelegt ist. Beispielfhaft können an dieser Stelle Video-Dateien oder vergleichbare anwendungsgebundene Dateien genannt werden.<sup>342</sup> Das Dateisystem und die Datenbanken sind über eine geeignete Schnittstelle für die Komponenten der Steuerungsschicht nutzbar.

Aufbauend auf der Beschreibung der Datenhaltung folgt im nachstehenden Kapitel eine Erläuterung der Steuerungsschicht.

---

338 Vgl. Lobe, Adrian: Wir sammeln eine Milliarde Seiten pro Woche, a. a. O., S. 13.

339 Vgl. Masanès, Julien: Web Archiving: Issues and Methods, a. a. O., S. 33 f.

340 Vgl. Ullrich, Dagmar: Bitstream Preservation, a. a. O., S. 166.

341 Vgl. Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 121 ff.

342 Vgl. Schicker, Edwin: Datenbanken und SQL: eine praxisorientierte Einführung mit Anwendungen in Oracle, SQL Server und MySQL, a. a. O., S. 225 ff.

### 5.2.4 Anwendungen der Steuerungsschicht

Die Steuerungsschicht ist durch den Einsatz von zwei lose gekoppelten Anwendungen charakterisiert (vgl. Abbildung 19 zur Einordnung der Anwendungen).<sup>343</sup>

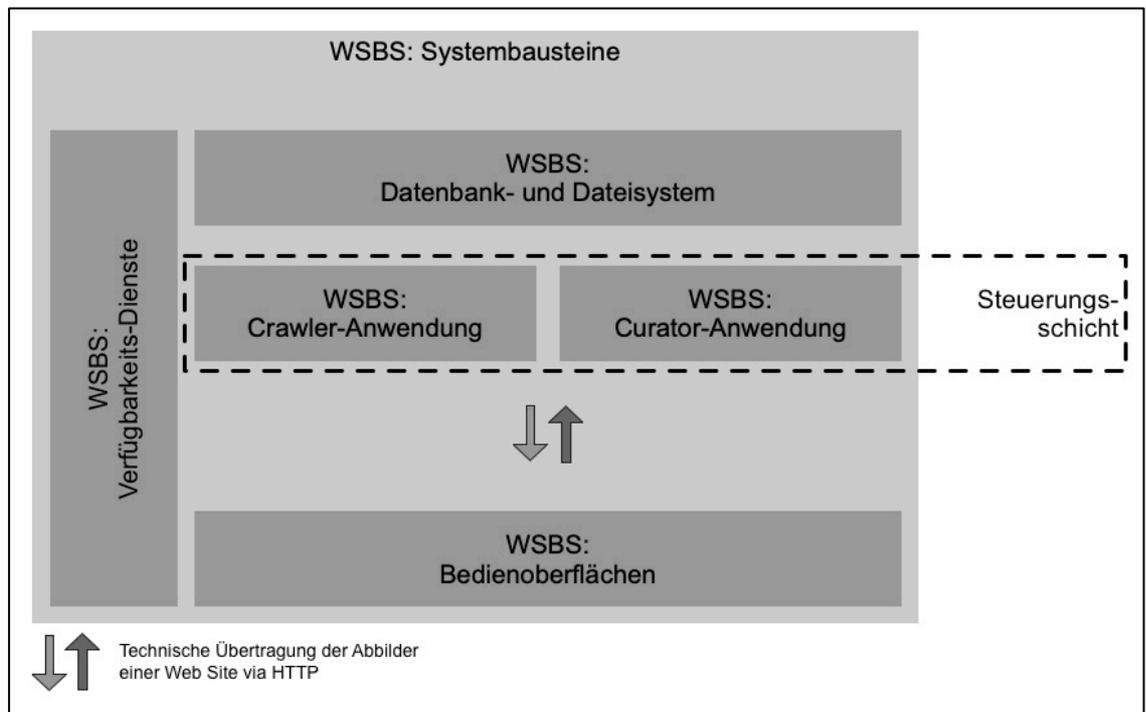


Abb. 19: Anwendungen zur Steuerung der Systemlösung

Ziel der Anwendung ist das Steuern sämtlicher Systemaufgaben. Die Systemaufgaben sind das Erfassen, Bewahren und Bereitstellen einer WSB. Um die Fehleranfälligkeit zu minimieren sowie die Performance und Wartbarkeit zu erhöhen, ist das Erfassen getrennt von der Bewahrung und Bereitstellung durchzuführen.<sup>344</sup> Die Geschäftslogik wird auf die Crawler- und Curator-Anwendung aufgeteilt.

Die Aufgabe der Crawler-Anwendung ist durch das Erfassen der WSB charakterisiert. Zentraler Bestandteil dieser Aufgabe ist das Analysieren von Veränderungen auf einer Web Site ebenso wie das Generieren von Abbildern zur Bewahrung. Sowohl Inhalte als auch Gestaltungselemente einer Web Site können mit der Anwendung erfasst und für die Nachwelt konserviert werden.

<sup>343</sup> Vgl. Bergsmann, Johannes; Unterauer, Markus: Requirements Engineering für die agile Softwareentwicklung: Methoden, Techniken und Strategien, a. a. O., S. 109.

<sup>344</sup> Vgl. z. B. Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 99 f.

Abbildung 20 ordnet der Crawler-Anwendung Aufgaben bzw. Module („Web Site analysieren“, „Abbilder generieren“) aus dem Strukturmodell zu.

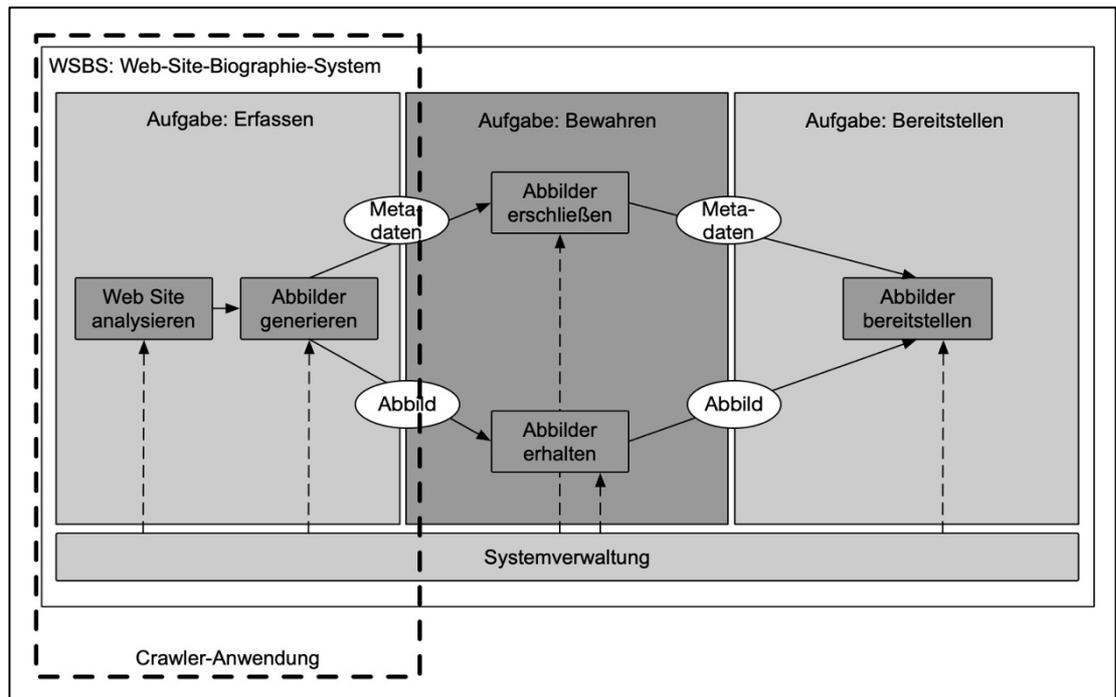


Abb. 20: Einordnung der Crawler-Anwendung im Strukturmodell

- Modul „Web Site analysieren“: Mit Hilfe der Funktionen dieses Moduls werden sämtliche Web-Seiten einer Web Site indexiert und ausgewertet. Dabei wird nicht nur textbasierter HTML-Quellcode auf Veränderungen überprüft, sondern auch verlinkte Dateien wie Downloads oder Bilder mit älteren Versionen in den Datenbanken bzw. Dateisystemen verglichen.
- Modul „Abbilder generieren“: Veränderungen werden in Abbildern festgehalten. Abbilder bestehen aus Kopien sämtlicher Dateien, die zum Darstellen einer Web Site verwendet werden. Metadaten zur Erschließung der Abbilder werden erzeugt und verarbeitet. Die generierten Metadaten beschreiben technische, administrative und strukturelle Gegebenheiten eines Abbildes. Beispielhaft sind an dieser Stelle Informationen wie Dateinamen, Dateipfade, technische Prüfsummen oder Zeitstempel zu nennen.<sup>345</sup>

<sup>345</sup> Vgl. Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 34 f. oder auch Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 66 ff.

Auf Basis der Metadaten werden die Abbilder für die Bewahrung und Bereitstellung zugänglich gemacht. Die Geschäftslogik für das Bewahren und Bereitstellen wird mit Hilfe der Curator-Anwendung implementiert.

Abbildung 21 hebt die Aufgaben bzw. Module der Curator-Anwendung („Abbilder erschließen“, „Abbilder erhalten“, „Abbilder bereitstellen“) im Strukturmodell hervor.

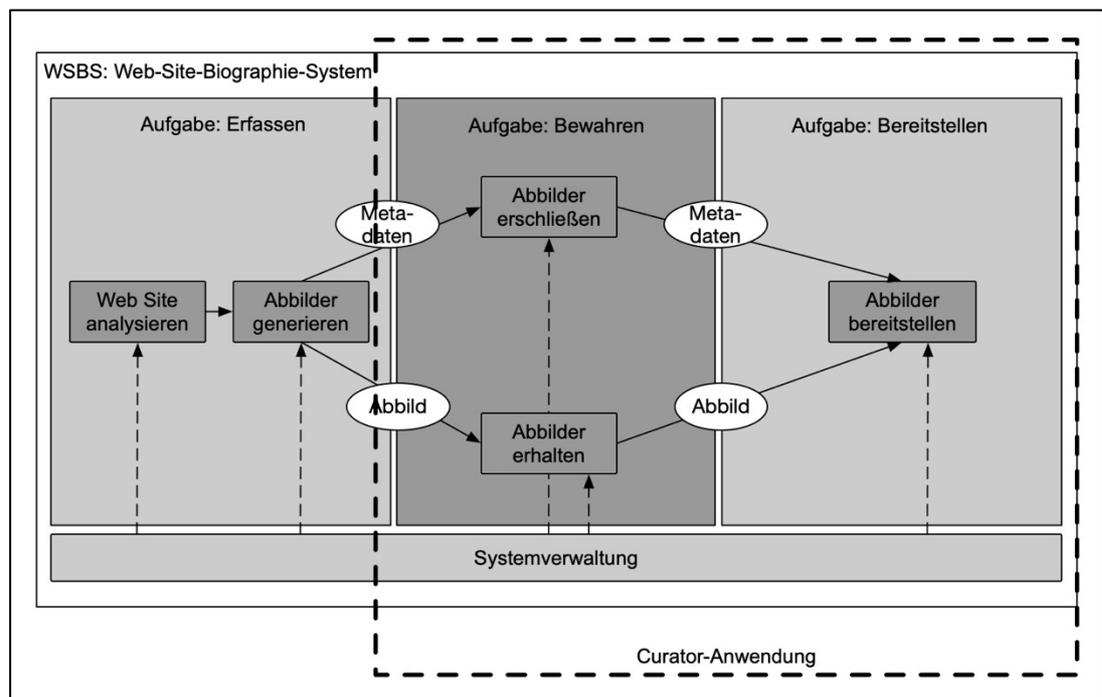


Abb. 21: Einordnung der Curator-Anwendung im Strukturmodell

Die Curator-Anwendung definiert, in welchen Dateiformaten Abbilder und Metadaten von der Crawler-Anwendung entgegengenommen werden. Abbilder werden mit ihren Metadaten in der Datenbank erschlossen und auf Basis von geeigneten Speichermedien für die Nachwelt konserviert.

- Modul „Abbilder erschließen“: Das Erschließen der Abbilder umfasst sämtliche Funktionen des Hinzufügens, Verwaltens und Prüfens der Abbilder, ebenso wie Funktionen für das Suchen und Interpretieren auf Basis ihrer Metadaten.<sup>346</sup>
- Modul „Abbilder erhalten“: Abbilder lassen sich zur Erhaltung mit einer geeigneten Funktionsmenge anpassen. Beispielhaft kann das Ersetzen kritischer Inhalte zur Wahrung der Rechtssicherheit der Organisation genannt werden, ebenso wie

<sup>346</sup> Vgl. Harvey, D. R.; Gillian, Oliver: Digital curation, a. a. O., S. 11 f. oder Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 10 f.

das Anpassen von Kontext-Informationen.<sup>347</sup> Darüber hinaus werden Mechanismen implementiert, die eine Migration veralteter Dateiformate ermöglichen, um die langfristige Verwendung der Abbilder zu gewährleisten.<sup>348</sup>

- Modul „Abbilder bereitstellen“: Mit diesem Modul, wird das Bereitstellen der Abbilder im Web realisiert. Es stellt notwendige Dateien aus der Datenbanklösung zusammen und überträgt diese mit Hilfe des HTTP.<sup>349</sup> Durch standardisierte Schnittstellen, die Inhalte und Gestalt der Abbilder in strukturierten Dateiformaten ausliefern, können skalierbare Oberflächen-Lösungen entwickelt werden.<sup>350</sup> Verändern sich Endgeräte zum Abrufen der WSB, lassen sich neue Bedienoberflächen ohne Beeinträchtigung produktiver Komponenten entwickeln.<sup>351</sup> Administratoren können bestimmen, welche Metadaten auf der Nutzungsoberfläche angezeigt bzw. berücksichtigt werden. Durch das Aktivieren bzw. Deaktivieren deskriptiver, administrativer, technischer oder struktureller Metadaten wird der Informationsgehalt einer WSB skaliert.

Nachdem im vorliegenden Kapitel die Anwendungen zur Steuerung der Geschäftslogik des WSBS erläutert wurden, beschäftigt sich das folgende Kapitel mit Bedienoberflächen, die Abbilder im WWW erlebbar machen.

---

347 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 50 f.

348 Vgl. Schumann, Natascha: Einführung in die digitale Langzeitarchivierung, a. a. O., S. 41 f.

349 Vgl. z. B. Saleh, Iman: Formalizing Data-Centric Web Services., Cham: Springer International Publishing 2015, S. 7.

350 Vgl. Patni, Sanjay: Pro RESTful APIs: Design, Build and Integrate with REST, JSON, XML and JAX-RS, a. a. O., S. 7 f.

351 Vgl. Doglio, Fernando: REST API Development with Node.js: Manage and Understand the Full Capabilities of Successful REST Development, 2018, S. 7 f.

### 5.2.5 Bedienoberflächen der Präsentationsschicht

Die Präsentation der Abbilder im WSBS wird durch Bedienoberflächen realisiert, die das Erfassen, Bewahren und Bereitstellen der Abbilder unterstützen (vgl. Abbildung 22 zur Einordnung in die Systembausteine des WSBS).

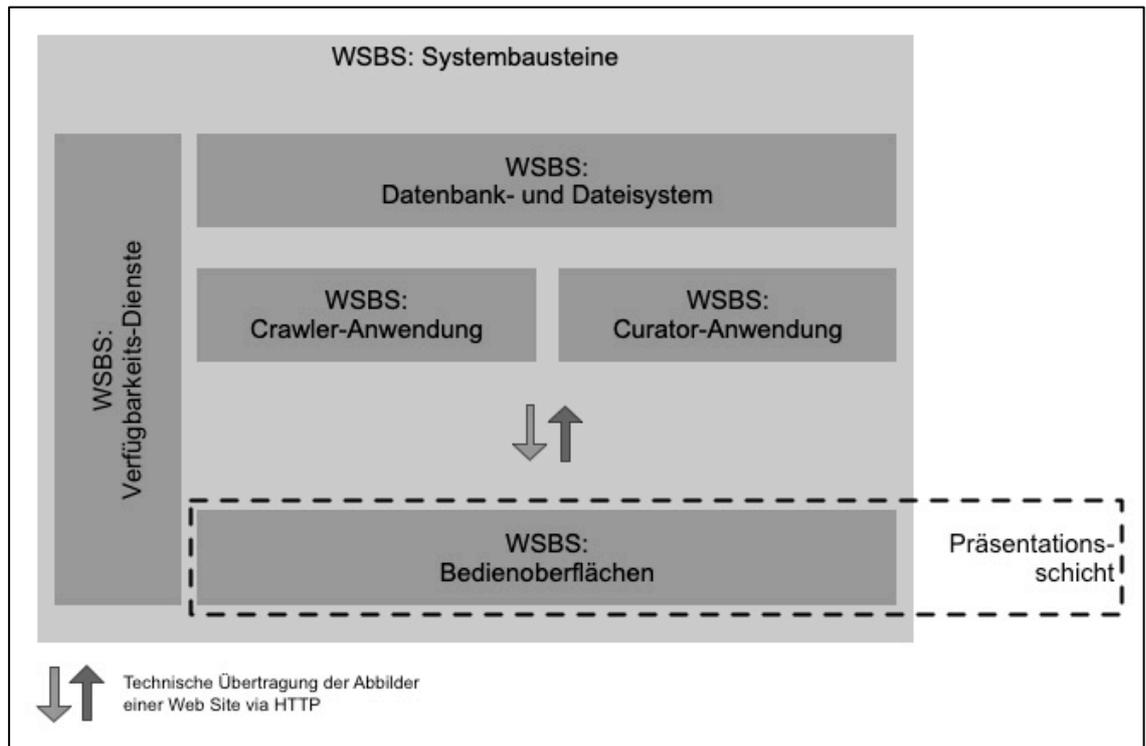


Abb. 22: Systembaustein der Präsentationsschicht im WSBS

Das zentrale Instrument zum Darstellen der Bedienoberflächen ist der Web Browser. Im vorliegenden Kapitel wird unter „Web Browser“ Software subsumiert, die es Nutzern ermöglicht, Dateien über das WWW abzurufen und darzustellen. Ein Web Browser kann sowohl auf klassischen Endgeräten wie Notebooks oder PC installiert sein, als auch auf einem Tablet, Smartphone oder vergleichbaren Endgeräten.<sup>352</sup>

Die Bedienoberflächen machen Administratoren das Erfassen und Verwalten der Abbilder auf einer Administrationsoberfläche möglich. (End-)Nutzer können dagegen die Lebensgeschichte einer Web Site über Nutzungsoberflächen abrufen.

Die Nutzungsoberflächen werden zum Recherchieren und Interpretieren der digitalen Vergangenheit verwendet und machen die WSB erlebbar. Um eine geeignete Bedien-

<sup>352</sup> Vgl. Chatfield, Tom; Freytag, Carl: 50 Schlüsselideen digitale Kultur, a. a. O., S. 28 ff.

oberfläche für Nutzer zu entwickeln, müssen Navigationselemente implementiert werden, die sowohl eine horizontale als auch eine vertikale Wanderung durch die Vergangenheit der Web Site ermöglichen.<sup>353</sup>

Eine horizontale Wanderung ist durch das Aufrufen von Zuständen zu unterschiedlichen Erfassungszeitpunkten charakterisiert. Die Erfassungszeitpunkte werden dabei z. B. über einen Zeitleiste oder kalendarische Übersichten zugänglich gemacht. Das Navigieren durch eine Menge von Abbildern zu *einem* definierten Erfassungszeitpunkt wird dagegen als vertikale Wanderung bezeichnet. Die vertikale Wanderung erfolgt klassischerweise durch das Klicken auf die archivierten Links in einem Abbild. Die Links verweisen auf weitere Abbilder, die dem gleichen Erfassungsprozess zugeordnet werden.

Abbildung 23 stellt eine Ausprägung der Nutzungsoberfläche zum Erleben einer WSB beispielhaft dar.<sup>354</sup>

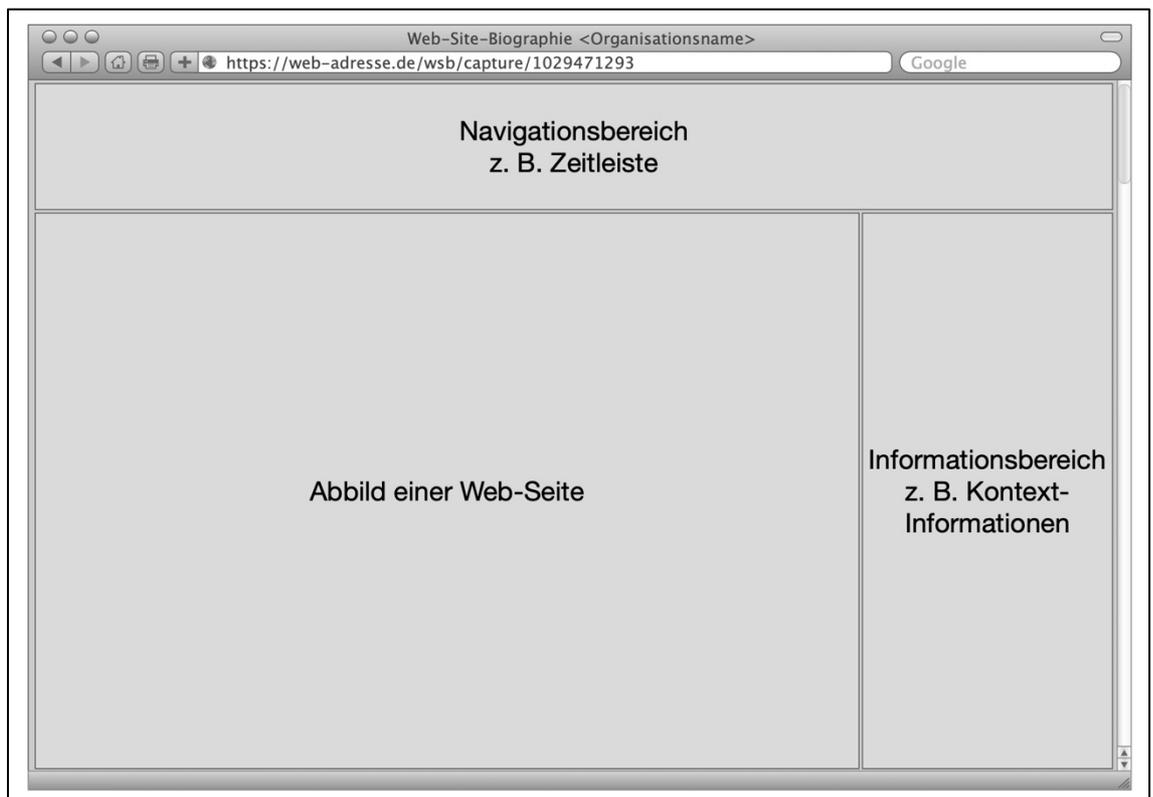


Abb. 23: Nutzungsoberfläche zum Erleben der WSB: Klassisch

353 Vgl. Ben-David, Anat; Huurdeman, Hugo: Web Archive Search as Research: Methodological and Theoretical Implications, in: Alexandria: The Journal of National and International Library and Information Issues, 25/2014, S. 95.

354 Vgl. Internet Archive (Hrsg.): Internet Archive Wayback Machine, a. a. O., 28.03.2017.

Die dargestellte Nutzungsoberfläche visualisiert eine typische Strukturierung innerhalb des Web Browsers.<sup>355</sup> In einem Navigationsbereich wird eine Zeitleisten-Navigation als Kopfzeile integriert. Das Abbild einer Web-Seite wird in Verbindung mit einem Informationsbereich abgerufen. Der Informationsbereich wird am Rand des Bildschirms dargestellt und das Abbild nimmt den größten Teil des zur Verfügung stehenden Platzes ein. Mit Hilfe der Zeitleiste können Abbilder zu unterschiedlichen Erfassungszeitpunkten ausgewählt und angezeigt werden.<sup>356</sup> Neben dieser klassischen Darstellung können aber auch komplexere Darstellungsformen zur Navigation realisiert werden. Nicht immer ist eine Zeitleistennavigation ausreichend.<sup>357</sup> Die Nutzungsoberfläche wird daher im Folgenden als Slide-Show-basierte Präsentationsumgebung skizziert.

Abbildung 24 stellt die Struktur der Nutzungsoberfläche mit Slide-Show beispielhaft dar.

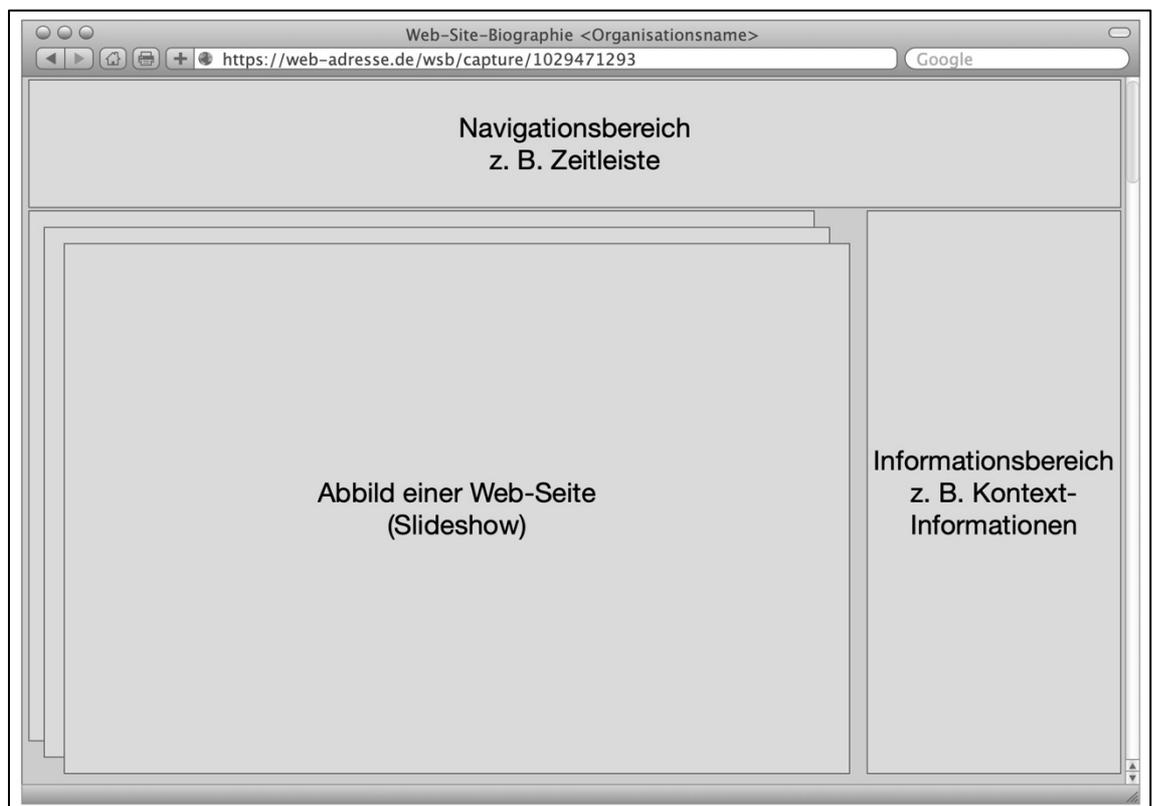


Abb. 24: Nutzungsoberfläche zum Erleben der WSB: Slide-Show

355 Vgl. Internet Archive (Hrsg.): Internet Archive Wayback Machine, a. a. O., 28.03.2017.

356 Vgl. für weitere Navigations- bzw. Darstellungsmöglichkeiten auch Meyer, Eric; Yasseri, Taha; Hale, Scott A.; Cowls, Josh; Schroeder, Ralph; Margetts, Helen: Analysing the UK web domain and exploring 15 years of UK universities on the web, in: The web as history: using web archives to understand the past and the present, Hrsg.: Brügger, Niels; Schroeder, Ralph, London: UCL Press 2017, S. 37.

357 Vgl. Grävemeyer, Arne: Mit der Zeitmaschine in die Altstadt, in: c't magazin für computer und technik, 15/2019, S. 128 ff.

Bei einer Slide-Show-basierten Darstellung werden die Abbilder einer Web Site zu einem Erfassungszeitpunkt versetzt und hintereinander dargestellt. Einzelne Abbilder können verschoben und vor- oder zurückliegende Abbilder geladen werden. Der Navigations- und Informationsbereich sind optional bzw. können ein- oder ausgeblendet werden.

Ergänzend zu der Slide-Show-basierten Darstellung ist denkbar, dass eine Nutzungsoberfläche das Auswählen und Präsentieren der Abbilder über eine drei-dimensionale Oberfläche ermöglicht.<sup>358</sup> Die Implementierung einer drei-dimensionalen Darstellungsumgebung ist mit rechenintensiven Prozessen im Web Browser verbunden. Eine entsprechende Implementierung ist bei Web-basierten Systemen jedoch eine beobachtbare Praxis.<sup>359</sup>

Abbildung 25 stellt die Bedienoberfläche zum Erleben der Abbilder in einer dreidimensionalen Oberflächenstruktur beispielhaft dar.

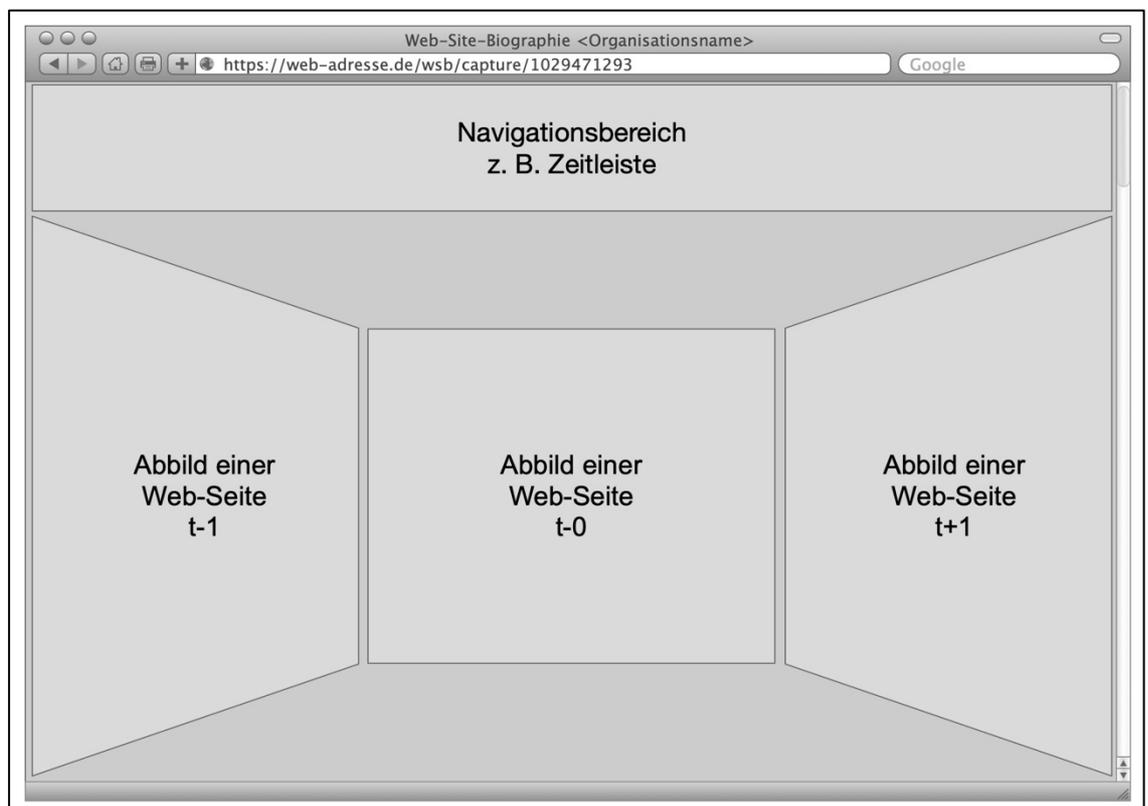


Abb. 25: Nutzungsoberfläche zum Erleben der WSB: Dreidimensional

358 Vgl. Meyer, Eric; Yasseri, Taha; Hale, Scott A.; Cowls, Josh; Schroeder, Ralph; Margetts, Helen: Analysing the UK web domain and exploring 15 years of UK universities on the web, a. a. O., S. 37.

359 Vgl. Herbig, Daniel: Google Stadia: Die Cloud-Gaming-Plattform, die Spieler über Youtube zocken lässt, Online im Internet: <https://www.heise.de/newsticker/meldung/Google-Stadia-Die-Cloud-Gaming-Plattform-die-Spieler-ueber-Youtube-zocken-laesst-4340325.html>, 10.04.2019.

Über die Zeitleiste können Erfassungszeitpunkte gewählt werden. Ein Abbild wird zum gewählten Erfassungszeitpunkt ( $t-0$ ) geladen und dargestellt. Zudem werden zeitlich vorliegende ( $t+1$ ) und zurückliegende ( $t-1$ ) Abbilder präsentiert. Die Abbilder können verglichen und zur detaillierten Betrachtung geöffnet werden. Ergänzend dazu ist denkbar, die Darstellung mit einer VR-Schnittstelle (Virtual-Reality-Schnittstelle) zu koppeln und eine Virtual-Reality-Umgebung zu ermöglichen.<sup>360</sup>

Nachdem im vorliegenden Abschnitt Skizzen für die Nutzungsoberflächen dargelegt wurden, widmet sich der folgende Abschnitt des Kapitels abschließend der Administrationsoberfläche. Die Administrationsoberfläche wird zum Verwalten der Systemfunktionen bzw. Module verwendet.

Abbildung 26 stellt eine klassische Struktur der Administrationsoberfläche zur Verwaltung der WSB dar.

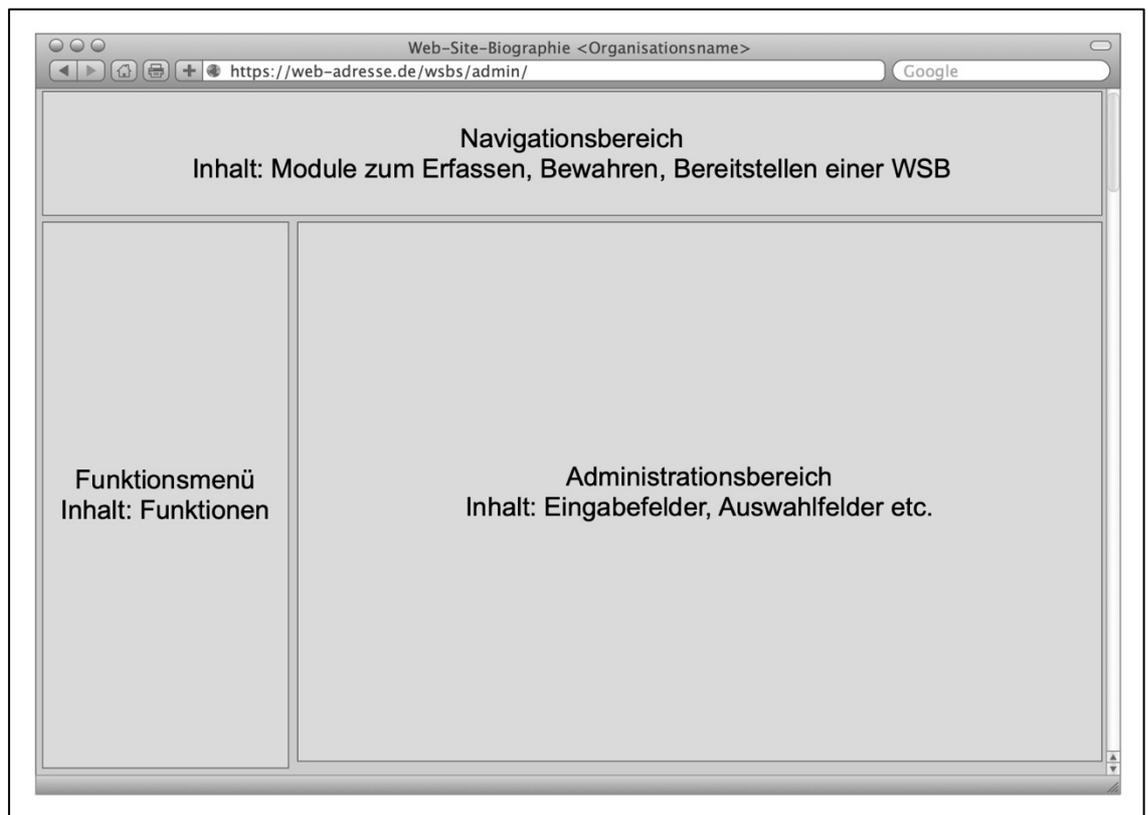


Abb. 26: Administrationsoberfläche

<sup>360</sup> Vgl. Krause, Dirk: VR im Web: Virtual-Reality-Web-Apps mit A-Frame programmieren, a. a. O., S. 116 oder z. B. Google LLC (Hrsg.): Building for VR on the Web, Online im Internet: <https://codelabs.developers.google.com/codelabs/webvr/index.html>, 19.04.2019.

In der Kopfzeile der Administrationsoberfläche ist ein Navigationsbereich implementiert, der die Auswahl von Modulen des IT-Systems ermöglicht. In Abhängigkeit des gewählten Moduls werden im Funktionsmenü Funktionen angezeigt, die das Konfigurieren und Nutzen des gewählten Moduls ermöglichen. Eine Auflistung aller verfügbarer Funktionen erfolgt im Laufe des Systementwurfs.

Aufbauend auf den skizzierten Bedienoberflächen werden im folgenden Kapitel zunächst Systembausteine skizziert, die für eine Sicherung und performante Nutzung der WSB bzw. des WSBS verantwortlich sind.

### 5.2.6 Verfügbarkeitsdienste

Die Verfügbarkeitsdienste umfassen Dienste, die im Regelfall nicht originär entwickelt werden müssen, sondern als ergänzende Bestandteile zur Absicherung des IT-Systems konfiguriert und integriert werden.

Abbildung 27 hebt die Verfügbarkeitsdienste in der schematischen Darstellung der Systembausteine hervor.

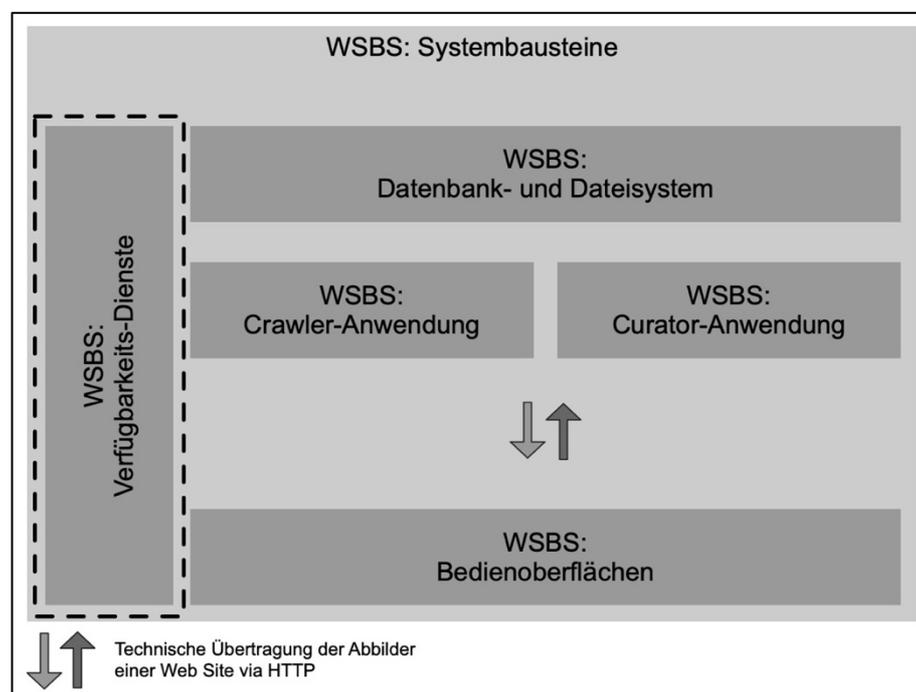


Abb. 27: Verfügbarkeitsdienste

Im vorliegenden Systementwurf werden die folgenden Verfügbarkeitsdienste zur Realisierung einer WSB berücksichtigt:<sup>361</sup>

- Firewall
- Proxy-Server
- Load Balancer
- Backup-System

Die Verfügbarkeitsdienste stellen sicher, dass die Systembausteine zum Erzeugen einer WSB verlässlich genutzt und vor ungewollten Systemangriffen oder Überlastungen geschützt werden. Der Aufgabenbereich der Firewall umfasst dabei das Filtern, Kontrollieren, Blockieren und Zulassen sämtlicher Kommunikationsversuche, die über das Internet mit dem Organisationsnetzwerk aufgebaut werden.<sup>362</sup> Die Firewall stellt eine abgesicherte Schnittstelle zwischen den originären Systembausteinen des WSBS und dem öffentlichen Internet dar.<sup>363</sup>

In Ergänzung zur Firewall ist der Einsatz eines Proxy-Servers möglich. Der Proxy-Server empfängt Anfragen und leitet diese an die geeigneten Zielsysteme weiter. Steuerungs- und Datenhaltungs-Komponenten können so z. B. ohne Auswirkungen auf den Produktivbetrieb (um-)konfiguriert und skaliert werden. Eine Erweiterung des Proxy-Servers stellt der Load Balancer dar. Klassischerweise wird ein Load Balancer für Server-Infrastrukturen eingesetzt, deren Ressourcen-Bedarfe zu Stoßzeiten nicht ohne Ausfallzeiten abgedeckt werden können. Eine Verteilung der Anfragen auf mehrere Server ist in diesem Fall möglich. Der Load Balancer regelt die Verteilung und legt fest, welche Anfragen an welchen Server weitergeleitet werden.<sup>364</sup> Für die Realisierung des WSBS-Prototyps ist die Integration eines Load Balancers optional. Sollten die System-Ressourcen bei der Abfrage einzelner Abbilder nicht mehr ausreichen, muss ein Load Balancer implementiert werden und eine Verteilung der Systembausteine auf mehrere Server erfolgen.

Das Backup-System kann als Erweiterung der Datenhaltungsschicht gesehen werden. Neben dem redundanten Speichern auf verteilten Datenbanken und Dateisystemen ist das

---

361 Vgl. zum Einsatz der genannten Verfügbarkeitsdienste im Unternehmenskontext Königs, Hans-Peter: IT-Risikomanagement mit System: praxisorientiertes Management von Informationssicherheits-, IT- und Cyber-Risiken, 5., überarbeitete und erweiterte Auflage, Wiesbaden: Springer Vieweg 2017, S. 193 f. oder Kappes, Martin: Netzwerk- und Datensicherheit: eine praktische Einführung, 2., überarb. und erw. Aufl., Wiesbaden: Springer Vieweg 2013, S. 183 ff.

362 Vgl. Kappes, Martin: Netzwerk- und Datensicherheit: eine praktische Einführung, a. a. O., S. 186.

363 Vgl. Liu, Alex X.; World Scientific (Firm): Firewall design and analysis, Singapore; Hackensack, N.J.: World Scientific Pub. Co. 2011, S. 1.

364 Vgl. zum Begriff „Load Balancer“ Dern, Gernot: Management von IT-Architekturen: Leitlinien für die Ausrichtung, Planung und Gestaltung von Informationssystemen, a. a. O., S. 327.

Integrieren einer Backup-Lösung wesentlich, um das Risiko eines Datenverlusts zu minimieren. Eine in der Praxis eingesetzte und etablierte Backup-Lösung ist z. B. der „Tivoli Storage Manager“ von IBM. Mit Hilfe dieser Backup-Software können Daten auf Bandspeicher in Backup-Umgebungen transferiert und bewahrt werden.<sup>365</sup>

Nachdem mit den Verfügbarkeitsdiensten der letzte Systembaustein beschrieben wurde, setzt sich das folgende Kapitel mit der Datenmodellierung auseinander.

### 5.3 Determinanten der Datenmodellierung: Abbilder und Metadaten

#### 5.3.1 Systematisierung der Datenmodellierung

Das Ziel des vorliegenden Kapitels ist der Entwurf von Datenstrukturen, die eine WSB auf Basis der zu verarbeitenden Daten beschreiben.<sup>366</sup> Grundlage des Datenmodells sind Abbilder und Metadaten, die für die Lebensgeschichte einer Web Site in einer Datenbanklösung gespeichert werden müssen.

Abbildung 28 hebt die Abbilder und Metadaten im Strukturmodell des WSBS hervor.

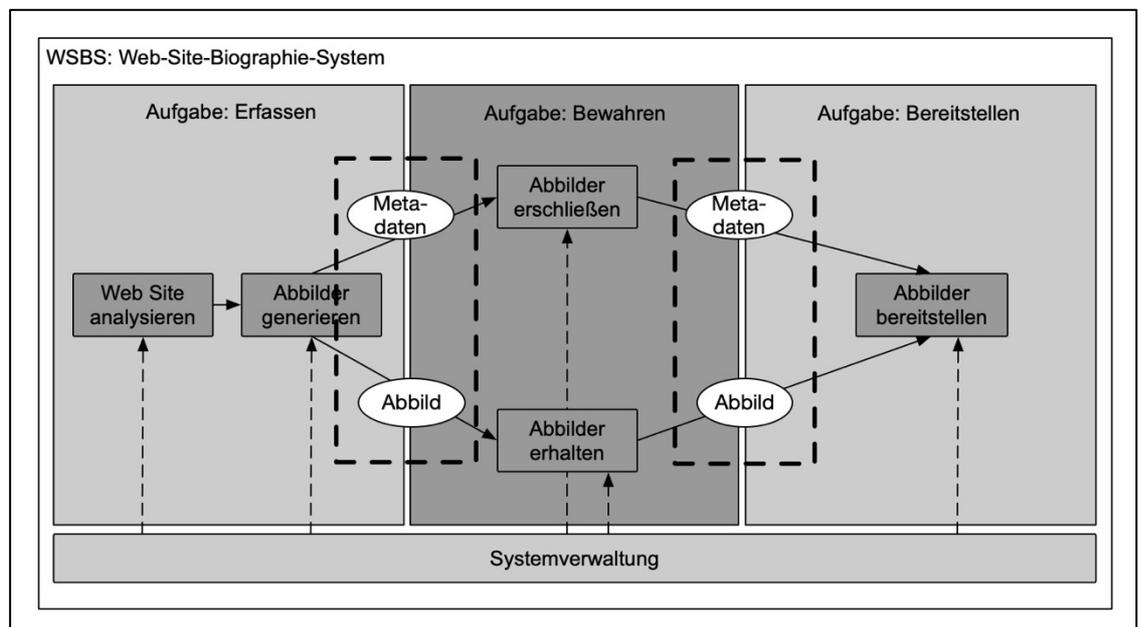


Abb. 28: Abbilder und Metadaten im Strukturmodell

365 Vgl. Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 121 ff.

366 Vgl. Gadatsch, Andreas: Datenmodellierung für Einsteiger, Wiesbaden: Springer Fachmedien Wiesbaden 2017, S. 1.

Zum Ableiten notwendiger Datenstrukturen wird eine konzeptionelle Beschreibung der Abbilder losgelöst von Datentypen und Datenbankformaten vorgenommen. Aufbauend auf der Beschreibung lassen sich technische Implementierungen für unterschiedlichste Datenbanksysteme ableiten.<sup>367</sup>

Damit die Abbilder einer Web Site langfristig interpretierbar bleiben, ist es notwendig, Metadaten der einzelnen Abbilder zu erfassen und zu pflegen.<sup>368</sup> Neben Kontext-Informationen, die das betriebsspezifische Umfeld der Abbilder beschreiben, müssen technische, administrative und strukturelle Informationen gesichert werden.<sup>369</sup>

Im folgenden Kapitel werden zunächst die Abbilder einer Web Site schematisch skizziert. Aufbauend auf dieser Skizze folgt die Beschreibung der Metadaten, die Abbilder langfristig interpretierbar machen.

### 5.3.2 Abbilder

Die Web Site einer Organisation ist aus technisch-konstruktiver Sicht als Konglomerat von miteinander verknüpften Web-Seiten zu verstehen.<sup>370</sup> Die technische Grundlage zum Erzeugen einer einzelnen Web-Seite ist eine HTML-Datei.<sup>371</sup>

Klassischerweise wird eine HTML-Datei nicht alleinstehend im Web verwendet, sondern mit weiteren Dateien verknüpft. Beispielhaft können an dieser Stelle PDF-Dateien oder Word-Dateien genannt werden, die auf einer Web-Seite als Download eingebunden und heruntergeladen werden können.<sup>372</sup>

Ergänzend zu den o. g. Dateien sind gestaltgebende Dateien omnipräsent. Neben Bild-Dateien gehören Style Sheets und Skript-Dateien zu etablierten Ressourcen auf nahezu jeder Web Site. Mit Style Sheets werden grafische Aufbereitungen (Farben, Formen,

---

367 Vgl. z. B. Gadatsch, Andreas: Datenmodellierung für Einsteiger, a. a. O., S. 2 f.

368 Vgl. Harvey, D. R.; Gillian, Oliver: Digital curation, a. a. O., S. 24 f.

369 Vgl. Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 36 f.

370 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 3.

371 Vgl. Schwickert, Axel C.: Web Site Engineering: Ökonomische Analyse und Entwicklungssystematik für eBusiness-Präsenzen, a. a. O., S. 98 ff.

372 Vgl. Krüger, Jörg Dennis; Kopp, Matthias: Web-Content managen: professioneller Einsatz von Content-Management-Systemen, a. a. O., S. 41.

Strukturen) der Inhalte vorgenommen. Skript-Dateien beinhalten dagegen Funktionen, die Interaktivität im Web ermöglichen.<sup>373</sup>

Damit ein originalgetreues Faksimile einer Web Site erfasst und bewahrt werden kann, müssen Kopien aller Web-Seiten erzeugt werden, die eine Organisation veröffentlicht. Typischerweise besteht eine einzelne Web-Seite jedoch nicht nur aus einer HTML-Datei, sondern ist mit weiteren Dateien verknüpft. Diese Dateien müssen ebenso kopiert und in der Datenbanklösung erschlossen werden. Das Konstrukt miteinander verknüpfter Dateikopien einer Web-Seite wird in der vorliegenden Arbeit als „Abbild“ bezeichnet.<sup>374</sup>

Abbildung 29 stellt den Verbund kopierter Dateien zur Realisierung eines einzelnen Abbildes beispielhaft und schematisch dar.

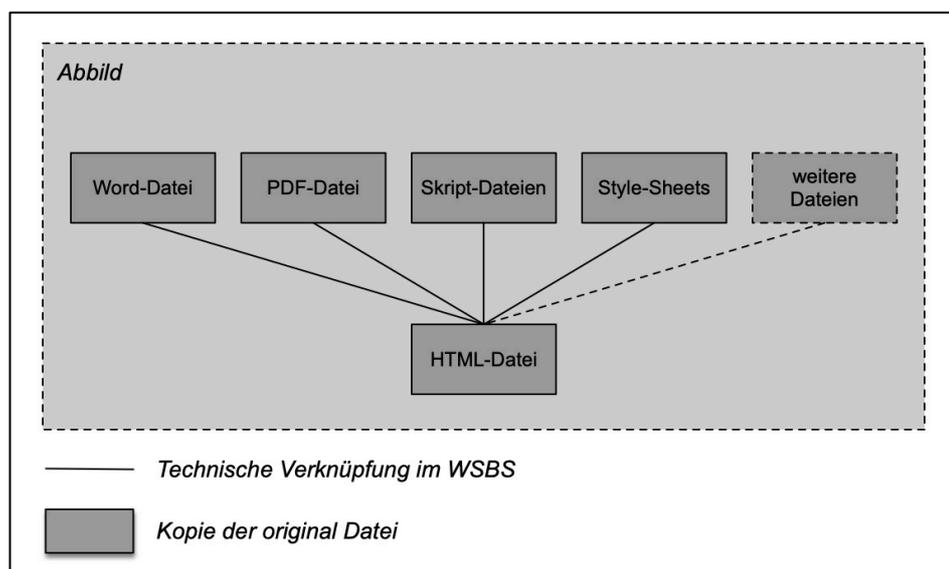


Abb. 29: Schematischer Aufbau eines Abbildes im WSBS

Neben dem Verknüpfen von mehreren Dateien miteinander, ist eine Mehrfachzuordnung der Dateien charakteristisch. Besonders bei Skript-Dateien oder Style Sheets findet häufig eine Mehrfachzuordnung zu unterschiedlichen HTML-Dateien statt. Diese Mehrfachzuordnung muss in geeigneter Art und Weise erhalten bleiben und in den Abbildern einer Web Site berücksichtigt werden.

373 Vgl. Collins, Mark J.: Pro HTML5 with CSS, JavaScript, and multimedia: complete website development and best practices, United States: Apress 2017, S. 29 ff.

374 Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 13 f.

Abbildung 30 stellt die Mehrfachzuordnung von einer Skript-Datei und einem Style Sheet beispielhaft dar.

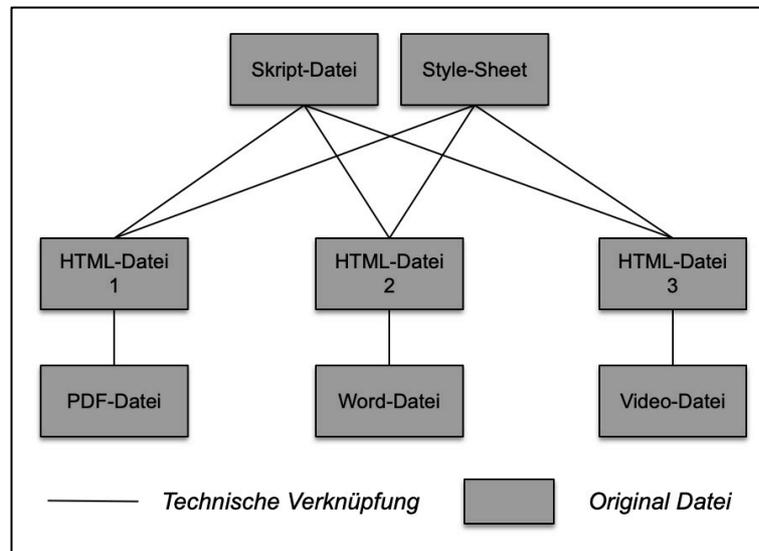


Abb. 30: Darstellung der Dateiverknüpfungen

Sollen die oben gezeigten Dateien und technischen Verknüpfungen auch in der WSB nutzbar sein, muss das Datenmodell konsequenterweise Datenbankstrukturen beinhalten, die diese Mehrfachzuordnungen berücksichtigen. Grundsätzlich kann in diesem Zusammenhang zwischen zwei Modellierungsweisen gewählt werden.

Zum einen können Datenbankstrukturen implementiert werden, die das Speichern isolierter Abbilder fordern. Dateien werden ohne Redundanzprüfung in den Datenbanken abgelegt und technische Verknüpfungen dokumentiert. Einzelne Abbilder werden auf diese Weise als unabhängige Artefakt erschlossen und können losgelöst voneinander reproduziert werden.

Abbildung 31 visualisiert schematisch das Konzept isolierter Abbilder (Abbild A, B, C).

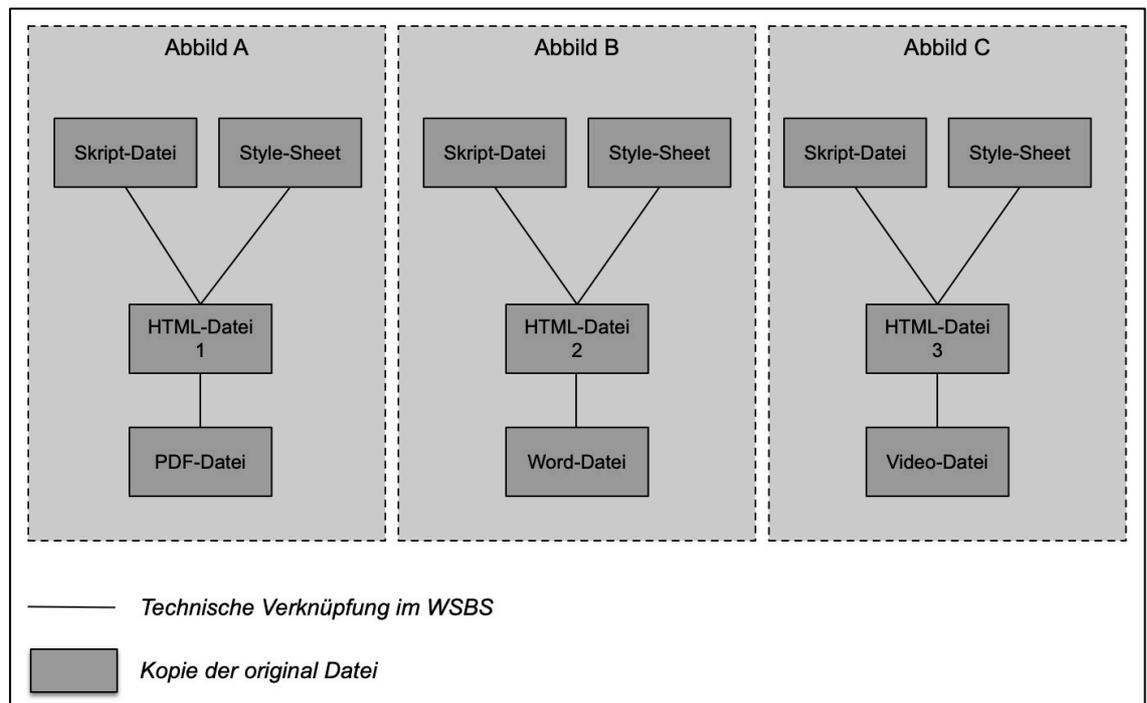


Abb. 31: Schematische Darstellung isolierter Abbilder

Ein Abbild besteht aus einer kopierten HTML-Datei (HTML-Datei 1, 2 oder 3) und einer eindeutig zugeordneten PDF- bzw. Word- oder Video-Datei. Darüber hinaus werden Kopien einer Skript-Datei und eines Style Sheets zugeordnet. Die Kopien der Skript-Datei und des Style Sheets sind im vorliegenden Fall jedoch nicht nur einmalig vorhanden, sondern werden für jedes Abbild individuell kopiert und zugeordnet. Es entstehen deckungsgleiche Kopien.

Zum anderen kann aber auch eine Mehrfachzuordnung im Rahmen der Datenmodellierung berücksichtigt werden. Um eine Mehrfachzuordnung von Dateien zu realisieren, muss mit Hilfe geeigneter Metadaten überprüft werden, ob Dateien zum Erzeugen eines Abbildes bereits in der Datenbank vorhanden sind. Neue Kopien von Dateien werden nur dann angefertigt, wenn eine identische Dateikopie noch nicht innerhalb der Speicherumgebung existiert.

Sind z. B. identische Versionen einer Skript-Datei und eines Style Sheets auf mehreren HTML-Dateien verlinkt, werden diese Dateien nur einmal erfasst und gespeichert. Das WSBS verknüpft die Kopie der Skript-Datei und die Kopie des Style Sheets automatisch mit allen relevanten Kopien der HTML-Dateien.

Die Mehrfachzuordnung von Dateien wird schematisch in Abbildung 32 dargestellt.

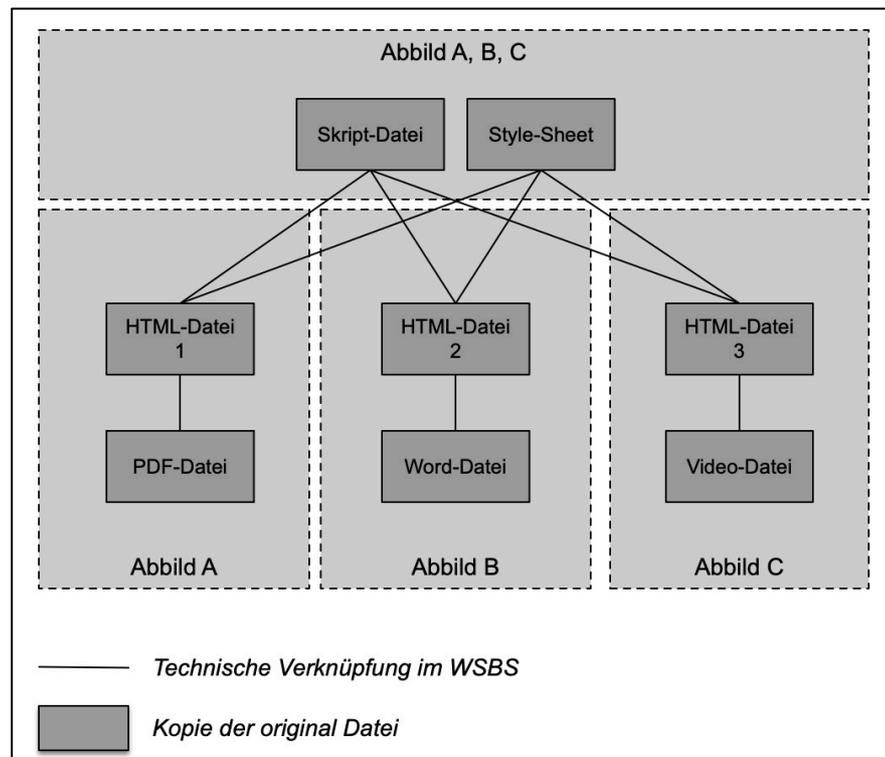


Abb. 32: Mehrfachzuordnung von Dateien im WSBS

Das vollständige Abbild einer Web-Seite „Abbild A“ besteht bei Mehrfachzuordnung aus originalgetreuen Kopien der HTML-Datei 1, der PDF-Datei sowie der Skript-Datei und des Style Sheets. Die Kopie der Skript-Datei und die Kopie des Style Sheets werden in diesem Fall einmal erzeugt und für „Abbild B“ und „Abbild C“ weiterverwendet. Eine solche Mehrfachzuordnung führt auf Datenhaltungsebene zu Ressourcen-Einsparungen und sollte aus Performance-Gründen berücksichtigt werden.

Nachdem im vorliegenden Kapitel Konzepte für die Realisierung der Abbilder auf Datenhaltungsebene beschrieben wurden, setzt sich das folgende Kapitel mit der Beschreibung von Metadaten auseinander. Metadaten sind notwendig, um die langfristige Interpretierbarkeit der Abbilder sicherzustellen.<sup>375</sup>

<sup>375</sup> Vgl. Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 66 f.

### 5.3.3 Metadaten

Im WSBS muss ein Set von Metadaten existieren, welches die Abbilder für Administratoren, Nutzer und Systembausteine interpretierbar macht.<sup>376</sup> Das Set an Metadaten wird als Bestandteil der WSB erfasst und auf Bedienoberflächen zur Verfügung gestellt.<sup>377</sup>

Metadaten werden gemeinhin als „Daten über Daten“ verstanden und umfassen im vorliegenden Systementwurf Informationen, die zur Interpretation der Abbilder notwendig sind.<sup>378</sup>

Für die Abbilder einer WSB lassen sich die Metadaten in vier Kategorien gliedern:<sup>379</sup>

- Deskriptive Metadaten
- Administrative Metadaten
- Technische Metadaten
- Strukturelle Metadaten

Deskriptive Metadaten beinhalten beschreibende Merkmale zur Dokumentation des Abbildes. Im Rahmen der Systemlösung umfassen die deskriptiven Metadaten Identifikationsmerkmale, den Namen bzw. Titel des Abbildes, Zeitstempel zur Dokumentation der Erstellung sowie die originale Web-Adresse der kopierten Datei. Darüber hinaus werden Informationen über unterschiedliche Versionen festgehalten ebenso wie sämtliche Informationen, die der Web Server über HTTP-Header zur Verfügung stellt. Durch das Sichern der Header-Informationen kann die technische Bereitstellung der original Web Site konserviert und adaptiert werden. Ergänzend dazu wird die technische und organisatorische

---

376 Vgl. Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 37.

377 Vgl. Aschenbrenner, Andreas; Rauber, Andreas: Mining Web Collections, a. a. O., S. 157 f.

378 Vgl. Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 36 f. oder Aschenbrenner, Andreas; Rauber, Andreas: Mining Web Collections, a. a. O., S. 157 f.

379 Vgl. Miller, Steven J.: Metadata for digital collections: a how-to-do-it manual, a. a. O., S. 10 ff. und Oermann, Andrea; Jäschke, Gerald; Dittmann, Jana: Vertrauenswürdige und abgesicherte Langzeitarchivierung multimedialer Inhalte, Frankfurt am Main: nestor - Kompetenznetzwerk Langzeitarchivierung c/o Deutsche Nationalbibliothek 2009, S. 25 und Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 37.

Infrastruktur zur Beschreibung der Betriebsspezifika erfasst. Neben einfachen Texten lassen sich z. B. Datei-Anhänge zur Infrastrukturbeschreibung hinzufügen. Beispielhaft sind hier Bilder oder Videos der betriebsrelevanten Hardware zu nennen.<sup>380</sup>

Administrative Metadaten umfassen dagegen Informationen über die Verwaltung und Verwendung einer Datei oder eines Dokuments.<sup>381</sup> In diesem Zusammenhang können Angaben über Zugriffsberechtigungen, Informationen über rechtliche Verpflichtungen oder Nutzungsstatistiken eines Abbildes angeführt werden. Darüber hinaus protokollieren administrative Metadaten die technischen Erhaltungsmaßnahmen, z. B. das erfolgreiche Durchführen von Migrationsprozessen.<sup>382</sup> Als Erweiterung der administrativen Metadaten beinhalten technische Metadaten Informationen wie Prüfsummen, Informationen über den Umfang und Zustand des Inhaltes (Anzahl an Zeichen im Quellcode), Angaben zum Dateityp (und implizit Angaben zur Software mit der die Datei erstellt wurde) sowie Informationen zur Verwendung.<sup>383</sup>

Zu den strukturellen Metadaten gehören Daten, die eine Einordnung des digitalen Artefakts in eine Gesamtmenge ermöglichen. Im WSBS werden auf Basis von strukturellen Metadaten technische Verknüpfungen zwischen Dateien abgeleitet. Beispielhaft kann eine Verknüpfung zwischen einer HTML- und einer Skript-Datei genannt werden.<sup>384</sup>

Unabhängig davon, ob es sich um deskriptive, administrative, technische oder strukturelle Metadaten handelt, wird im vorliegenden Systementwurf eine weitere Differenzierung in automatisiert und manuell erstellte Metadaten vorgenommen.

Automatisiert erstellte Metadaten lassen sich bereits während der Erfassung generieren und einem Abbild zuordnen. Beispielhaft sind hier Dateitypen oder Identifizierungsmerkmale zu nennen. Manuell erstellte Metadaten müssen dagegen durch die Betreiber bzw. Administratoren des WSBS eingepflegt werden. Beispielhaft können hier deskriptive Metadaten zur Infrastruktur genannt werden.

---

380 Vgl. Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 37.

381 Vgl. Harvey, D. R.; Gillian, Oliver: Digital curation, a. a. O., S. 67.

382 Vgl. Miller, Steven J.: Metadata for digital collections: a how-to-do-it manual, a. a. O., S. 10 ff.

383 Vgl. Oermann, Andrea; Jäschke, Gerald; Dittmann, Jana: Vertrauenswürdige und abgesicherte Langzeitarchivierung multimedialer Inhalte, a. a. O., S. 25.

384 Vgl. Harvey, D. R.; Gillian, Oliver: Digital curation, a. a. O., S. 65 ff. oder Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 66 f.

Tabelle 1 stellt relevante Metadaten strukturiert nach ihrer Erstellung (automatisiert/manuell) und Verwendung (deskriptiv, administrativ, technisch, strukturell) dar.<sup>385</sup>

	Deskriptive Metadaten	Administrative Metadaten	Technische Metadaten	Strukturelle Metadaten
Automatisierte Erstellung	<ul style="list-style-type: none"> <li>– Identifikationsmerkmale</li> <li>– Dateiname bzw. Titel</li> <li>– Web-Adresse</li> <li>– Zeitstempel</li> <li>– Version</li> <li>– HTTP-Header</li> </ul>	<ul style="list-style-type: none"> <li>– Abrufstatistiken</li> <li>– Erhaltungsmaßnahmen</li> </ul>	<ul style="list-style-type: none"> <li>– Prüfsumme</li> <li>– Dateityp</li> <li>– Codierung</li> <li>– Dateigröße</li> </ul>	<ul style="list-style-type: none"> <li>– Verknüpfte Dateien</li> <li>– Vorgänger-Nachfolger-Beziehungen</li> </ul>
Manuelle Erstellung	<ul style="list-style-type: none"> <li>– Technische Infrastruktur</li> <li>– Organisatorische Infrastruktur</li> </ul>	<ul style="list-style-type: none"> <li>– Zugriffsberechtigungen</li> <li>– Rechtliche Informationen</li> </ul>		

Tab. 1: Metadaten der Abbilder im WSBS

Eine Differenzierung in automatisiert und manuell erstellte Metadaten wird im vorliegenden Systementwurf durchgeführt, um die Entwicklung von Funktionen bestmöglich zu unterstützen. Die Crawler-Anwendung muss Programm-Code implementieren, der die automatisiert zu erfassenden Metadaten eigenständig ableitet und speichert. Die Curator-Anwendung muss zudem Eingabemasken bereitstellen, die das Erschließen bzw. Editieren der manuell zu erstellenden Metadaten ermöglichen.

Grundsätzlich sind bei der Modellierung auf Datenbankebene Kardinalitäten zu berücksichtigen. Vergleichbar mit Bildern oder Büchern, die in Archiven aufbewahrt und nach Autor, Ersteller oder Zeitepochen zusammengefasst werden, können Metadaten definiert werden, die für ein Menge an Abbildern oder ein spezifisches Abbild Gültigkeit besitzen. Beispielhafte Metadaten, die für eine Menge von Abbildern verwendet werden, sind betriebsspezifische Informationen, die Abbilder in einem organisatorisch-technischen Kontext interpretierbar machen. Bei der Implementierung geeigneter Datenbankstrukturen ist auch diese Mehrfachzuordnung der Metadaten zu berücksichtigen.<sup>386</sup>

Aufbauend auf der Modellierung von Abbildern und Metadaten folgt die Auseinandersetzung mit der Implementierung des Datenmodells.

<sup>385</sup> Vgl. Schrimpf, Sabine: Das OAIS-Modell für die Langzeitarchivierung: Anwendung der ISO 14721 in Bibliotheken und Archiven, a. a. O., S. 128.

<sup>386</sup> Vgl. Miller, Steven J.: Metadata for digital collections: a how-to-do-it manual, a. a. O., S. 33 f.

### 5.3.4 Zur Implementierung eines Datenmodells

Die Datenhaltung für eine WSB ist durch den Einsatz von Datenbanken umzusetzen. Das vorliegende Kapitel erläutert technische Realisierungen dieser Datenbanken.

In Organisationen werden Datenbanken im Regelfall eingesetzt, um eine zentrale, dauerhafte und strukturierte Sicherung von Unternehmens- bzw. Geschäftsprozessdaten zu ermöglichen. Die Datenbanken werden mit einem Datenbank-Management-System (DBMS) erstellt und durch technische Anwendungen oder Mitarbeiter befüllt. Der Zusammenschluss aus DBMS und Datenbanken (DB) wird allgemein hin als Datenbanksystem bezeichnet.<sup>387</sup> Bei Datenbanksystemen kann grundsätzlich zwischen relationalen und nicht-relationalen Systemen unterscheiden werden.<sup>388</sup>

Bei relationalen Datenbanksystemen werden Daten in Tabellen gespeichert und bereitgestellt. Das Datenmodell einer relationalen Datenbank berücksichtigt funktionale Abhängigkeiten und eliminiert Redundanzen durch Normalisierung. Mit der Einhaltung von Normalformen wird eine konsistente Datenhaltung realisiert. In einem Schema zur Beschreibung der Tabellen werden Spalten und Zeilen sowie deren Merkmalsausprägungen wie Datentypen und Identifikationsschlüssel definiert.<sup>389</sup> Zur Erstellung, Abfrage und Manipulation von Datenbankinhalten existieren standardisierte Datenbanksprachen.<sup>390</sup>

Bis heute sind relationale Datenbanksysteme aus der IT-Landschaft von Unternehmen bzw. Organisationen nicht wegzudenken und als Standard in der Datenhaltung und Datenbereitstellung etabliert.<sup>391</sup> Bei hohen Verarbeitungsgeschwindigkeiten, heterogenen Daten und einer hohen Skalierbarkeit ermöglichen relationale Implementierungen jedoch nicht immer eine effiziente Verwaltung. Nicht-relationale Datenbanksysteme stellen in diesem Zusammenhang eine Alternative dar.<sup>392</sup>

Nicht-relationale Datenbanksysteme sind häufig durch eine verteilte Systemarchitektur charakterisiert und speichern Daten losgelöst von fixen Schemata.<sup>393</sup> Neben sogenannten

---

387 Vgl. Meier, Andreas; Kaufmann, Michael: SQL- & NoSQL-Datenbanken, 8., überarbeitete und erweiterte Auflage, Berlin Heidelberg: Springer Vieweg 2016, S. 9 ff.

388 Vgl. z. B. MongoDB, Inc.(Hrsg.): Relational Vs Non Relational Database, Online im Internet: <https://www.mongodb.com/scale/relational-vs-non-relational-database>, 22.04.2019.

389 Vgl. Meier, Andreas: Werkzeuge der digitalen Wirtschaft: Big Data, NoSQL & Co Eine Einführung in relationale und nicht-relationale Datenbanken, Wiesbaden: Springer Vieweg 2018, S. 9 f.

390 Vgl. Meier, Andreas: Werkzeuge der digitalen Wirtschaft: Big Data, NoSQL & Co Eine Einführung in relationale und nicht-relationale Datenbanken, a. a. O., S. 9 oder Meier, Andreas; Kaufmann, Michael: SQL- & NoSQL-Datenbanken, a. a. O., S. 6 ff.

391 Vgl. Meier, Andreas; Kaufmann, Michael: SQL- & NoSQL-Datenbanken, a. a. O., S. 187.

392 Vgl. Meier, Andreas; Kaufmann, Michael: SQL- & NoSQL-Datenbanken, a. a. O., S. 11.

393 Vgl. Sadalage, Pramod J.; Fowler, Martin: NoSQL distilled: a brief guide to the emerging world of polyglot persistence, Upper Saddle River, NJ: Addison-Wesley 2013, S. 10.

„Key-Value“-Paaren oder Graphen-Datenbanken werden Dokumenten-Speicher zum persistenten Ablegen von Daten verwendet. Eine Strukturierung der Daten kann bei nicht-relationalen Implementierungen flexibel erweitert werden und über die Zeit variieren.<sup>394</sup>

Im Gegensatz zu relationalen Datenbanken können in einem Dokumenten-Speicher unterschiedlich strukturierte Dokumente abgelegt und verwendet werden. Durch das Aussparen einer Fixierung auf verbindliche Schemata wird eine vergleichsweise große Flexibilität erreicht. Diese Flexibilität macht nicht-relationale Datenbanksysteme besonders für die Verarbeitung heterogener Datenmengen attraktiv. Mit Bezug zur Langzeiterhaltung in einem WSBS bieten nicht-relationale Datenbanksysteme die Möglichkeit, Anpassungen an den Datenmodellen der Abbilder vorzunehmen, ohne Strukturen älterer Datensätze zu beeinträchtigen. Abbilder mit unterschiedlichen Strukturierungen können ohne Probleme in einem Dokumenten-Speicher abgelegt werden.<sup>395</sup> Zudem kann die Verarbeitungszeit von verteilten und großen Datenmengen als Vorteil nicht-relationaler Datenbanksysteme angeführt werden.<sup>396</sup>

Da zukünftige Anforderungen an die Datenhaltung schwer vorherzusehen sind und in der Fachliteratur keine eindeutigen Empfehlungen zur langfristigen Bewahrung in relationalen oder nicht-relationalen Systemen vorgenommen werden, soll zunächst keine Implementierungsform ausgeschlossen werden.

Um die allgemeine Gültigkeit des Systementwurfs zu wahren, werden die Abbilder und Metadaten losgelöst von einer technischen Realisierung entworfen. Die Implementierung kann im Rahmen relationaler oder nicht-relationaler Systeme konkretisiert werden und wird im Folgenden nicht für ein ausgewähltes Datenbanksystem spezifiziert.

---

394 Vgl. Copeland, Rick: MongoDB applied design patterns, Sebastopol, CA: O'Reilly Media, Inc 2013, S. 6 ff. oder Meier, Andreas; Kaufmann, Michael: SQL- & NoSQL-Datenbanken, a. a. O., S. 18 f.

395 Vgl. Copeland, Rick: MongoDB applied design patterns, a. a. O., S. 8 f.

396 Vgl. Meier, Andreas: Werkzeuge der digitalen Wirtschaft: Big Data, NoSQL & Co Eine Einführung in relationale und nicht-relationale Datenbanken, a. a. O., S. 43 ff.

## 5.4 Module und Funktionen

### 5.4.1 Systematisierung der Module und Funktionen

Aufbauend auf der Datenmodellierung werden im vorliegenden Kapitel Systemfunktionen skizziert. Die Systemfunktionen stehen auf der Bedienoberfläche zur Verfügung, um eine WSB erlebbar zu machen. Grundlage für das Bestimmen der Systemfunktionen bilden die fachlichen Anforderungen. Ziel der fachlichen Anforderungen ist das Beschreiben eines Zielsystems auf Basis seiner Funktionen.<sup>397</sup>

Für den Systementwurf werden (System-)Funktionen aus den fachlichen Anforderungen abgeleitet und in Modulen kumuliert.<sup>398</sup>

Abbildung 33 stellt die Module zunächst überblicksartig dar.<sup>399</sup>

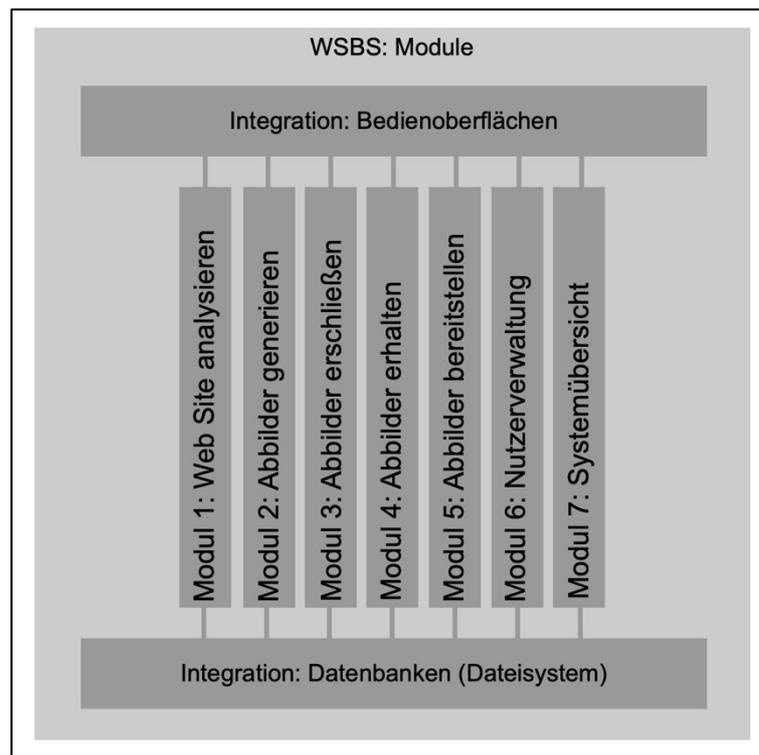


Abb. 33: Module des WSBS

397 Vgl. Ludewig, Jochen; Lichter, Horst: Software Engineering: Grundlagen, Menschen, Prozesse, Techniken, a. a. O., S. 369 ff.

398 Vgl. Balzert, Helmut: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, a. a. O., S. 142.

399 Vgl. Gruhn, Volker; Striemer, Rüdiger: The essence of software engineering, Schweiz: Springer 2018, S. 175.

Die Module sind durch funktionale Einheit, Unabhängigkeit und einen überschaubaren Funktionsumfang ausgezeichnet.<sup>400</sup> Sie übernehmen klar definierte Aufgaben.<sup>401</sup>

Funktionen für das Erfassen werden in den Modulen „Modul 1“ und „Modul 2“ zusammengefasst. Funktionen, die das Bewahren ermöglichen, werden mit „Modul 3“ und „Modul 4“ skizziert. Das Bereitstellen wird durch Funktionen des Moduls „Modul 5“ erläutert. Darüber hinaus existieren die Module „Modul 6“ und „Modul 7“ als systemübergreifende Komponenten, die eine Zugriffs- und Prozesskontrolle ermöglichen.

#### 5.4.2 Erfassen

Auf den folgenden Seiten werden die Module zum Erfassen der Abbilder skizziert und erläutert. Abbildung 34 grenzt die Module in der Modulübersicht ab.

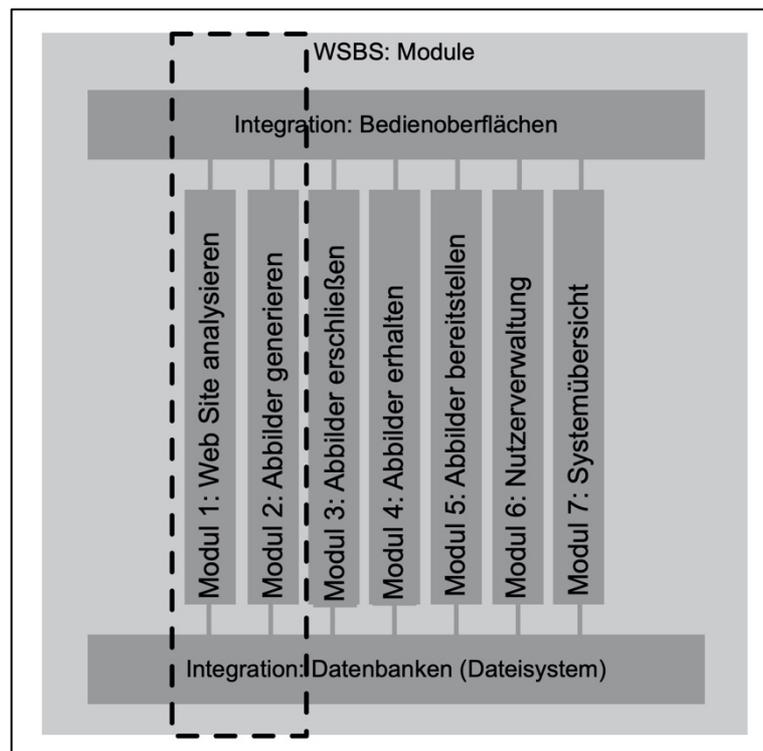


Abb. 34: Module zum Erfassen im WSBS

400 Vgl. Balzert, Helmut: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, a. a. O., S. 41.

401 Vgl. Balzert, Helmut: Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, a. a. O., S. 143 f.

Die Module zum Erfassen einer WSB gliedern sich in das „Web Site analysieren“ und „Abbilder generieren.“ Aus analytischen Gründen findet eine Untergliederung der Funktionen in Sub-Module statt.

In den Sub-Modulen „Adressraumverwaltung“, „Erfassungsbereiche und Frequenzen“ sowie „Abgesicherte Bereiche“ werden einzelne Funktionen zusammengefasst und im nachstehenden Textabschnitt erläutert.

Abbildung 35 stellt die (Sub-)Module zum Analysieren einer Web Site in einem Funktionsbaum dar.

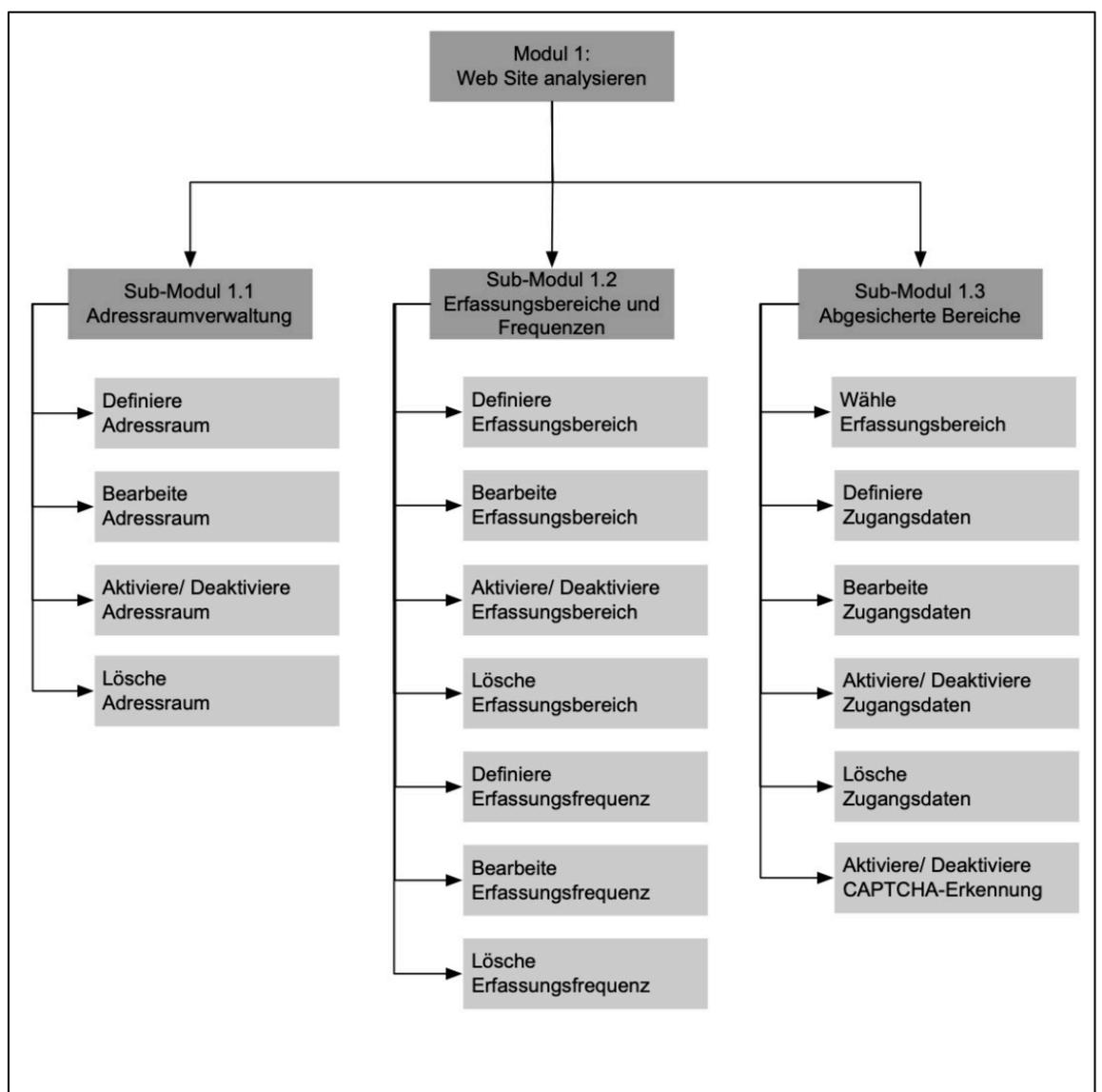


Abb. 35: Module und Funktionen zur Erfassung im WSBS (1/2)

Zum „Web Site analysieren“ werden im Folgenden die Sub-Module aus Abbildung 35 aufgelistet und Funktionen detaillierter beschrieben:

- „Adressraumverwaltung“: Zur Verwaltung des Adressraums werden Funktionen bereitgestellt, die auf Basis einer Web-Adresse den Einsatzbereich des WSBS abgrenzen und editierbar machen. Der Adressraum wird durch die Domäne der organisationalen Web Site (z. B. „uni-giessen.de“) bestimmt und um Protokoll-Angaben (HTTP vs. HTTPS) erweitert. Adressräume können aktiviert und deaktiviert werden.
- „Erfassungsbereiche und Frequenzen“: Aufbauend auf der Definition des Adressraums werden Erfassungsbereiche und Frequenzen definiert. Die Erfassungsbereiche werden als relative Web-Adressen (z. B. „/“, „/news“) für einen Adressraum angegeben. Erfassungsfrequenzen werden pro Erfassungsbereich festgelegt. Sie bestimmen, in welchen Intervallen Web-Seiten innerhalb eines oder mehrerer Erfassungsbereiche analysiert und kopiert werden. Der Erfassungstyp bestimmt zudem, ob Abbilder vollständig oder differenziert pro Erfassungsbereich erfasst werden. Ein vollständiges Erfassen setzt keine Analyse der Veränderungen voraus, sondern resultiert in einem umfassenden Abbild. Der Status quo einer Web Site wird kopiert und an die Speicherumgebung des WSBS übergeben. Ein differenziertes Erfassen ermöglicht das exklusive Kopieren von Web-Seiten, auf denen Veränderungen feststellbar sind. Erfassungsprozesse können mit Zeit und Datumsangaben geplant oder unverzüglich gestartet werden.
- „Abgesicherte Bereiche“: Zum Erfassen abgesicherter Bereiche wird ein Erfassungsbereich ausgewählt und Zugangsdaten (Benutzername und Passwort) zugeordnet. Auf diese Weise wird dem System mitgeteilt, dass für den gewählten Erfassungsbereich eine Benutzername-Passwort-Kombination zum Abfragen der Inhalte und Gestaltungselemente notwendig ist. Optional ist das Aktivieren oder Deaktivieren alternativer Sicherheitsabfragen wie einer CAPTCHA-Erkennung, die zum automatisierten Ausfüllen der Sicherheitsabfrage verwendet wird.

Ergänzend zu den Funktionsmengen in Abbildung 34 visualisiert der Funktionsbaum in Abbildung 35 (Sub-)Module und Funktionen des Moduls „Abbilder generieren“. Die

Funktionen der Sub-Module ermöglichen eine „Dateiverwaltung“, „Negativlisten“ und „Positivlisten“.

Abbildung 36 stellt die (Sub-)Module zum Generieren der Abbilder in einem Funktionsbaum überblicksartig dar.

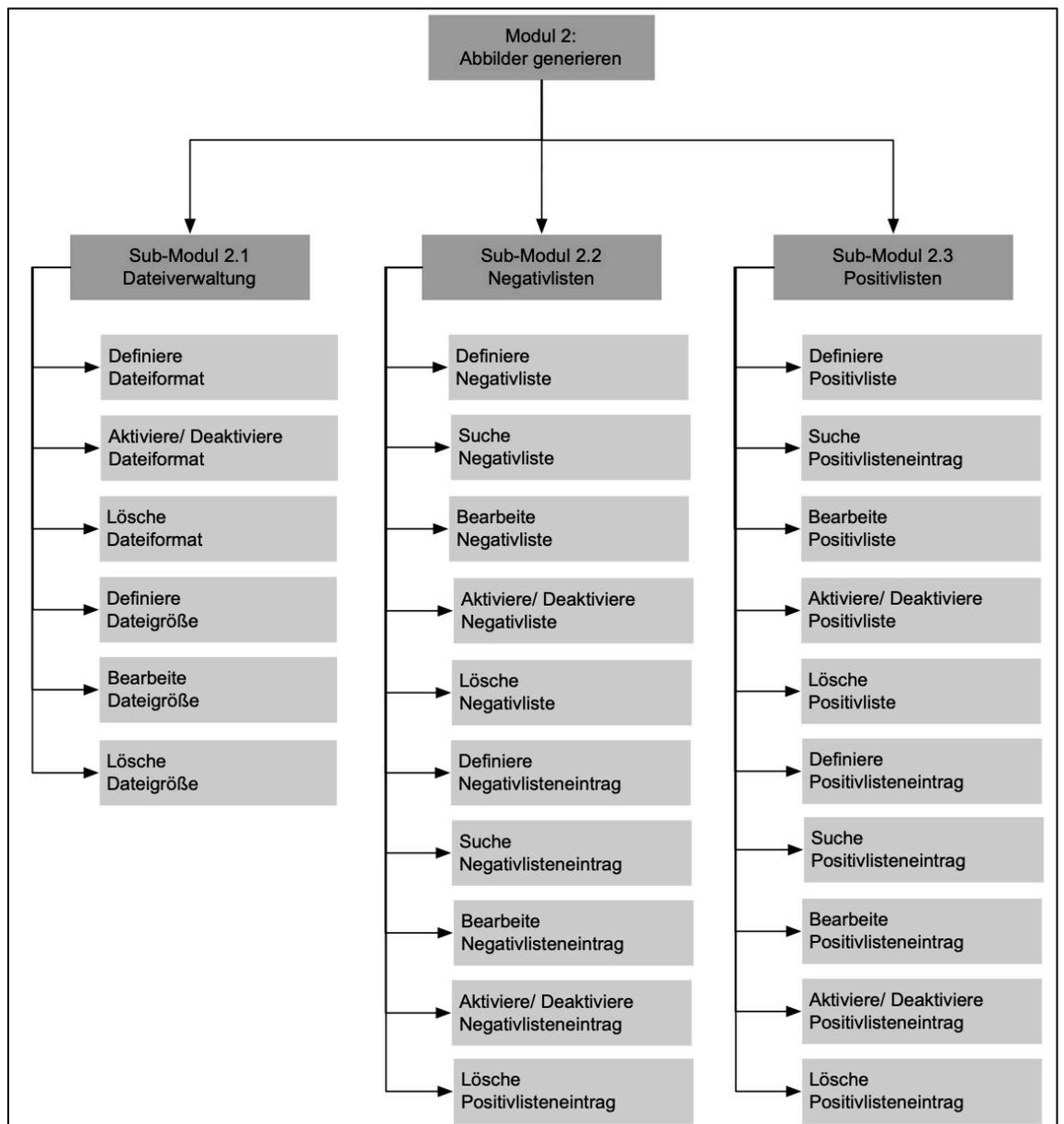


Abb. 36: Module und Funktion zur Erfassung im WSBS (2/2)

Zum „Abbilder generieren“ werden im Folgenden die Sub-Module aufgelistet und Funktionen beschrieben:

- „Dateiverwaltung“: Die Dateiverwaltung ermöglicht das Festlegen von Dateiformaten und Dateigrößen, die bei der Erstellung eines Abbildes berücksichtigt werden. Nur Dateien, die entsprechende Dateiformate und Dateigrößen (z. B. max.

100 MB) aufweisen, werden von der original Web Site kopiert und in die Speicherumgebung des WSBS überführt. Ein nachträgliches Bearbeiten der Liste ist möglich. Es können neue Dateiformate konfiguriert und veraltete Dateiformate deaktiviert werden. Auch ein Löschen der Dateiformate und Dateigrößen aus der Konfiguration ist möglich.

- „Negativlisten: Das proaktive Exkludieren von kritischen Inhalten erfolgt durch das Erstellen von Negativlisten. Eine Negativliste beinhaltet Web-Adressen, die auf Inhalte im Adressraum der Organisation verweisen. In der Liste können z. B. Web-Adressen als Negativlisteneintrag gepflegt werden, die auf Fotos von Mitarbeitern verweisen, die aus rechtlichen Gründen von der Erfassung auszuschließen sind. Das Suchen, Bearbeiten und Löschen der Negativlisten und Negativlisteneinträge ist möglich.
- „Positivlisten“: Das Pendant zu Negativlisten sind Positivlisten. In einer Positivliste werden Web-Adressen definiert, bearbeitet oder gelöscht, die auf Inhalte externer Anbieter verweisen. Im Gegensatz zur Negativliste verweist ein Eintrag in einer Positivliste im Regelfall auf Web-Adressen, die außerhalb des Adressraums der Organisation liegen. Die Positivliste legitimiert externe Inhaltsdateien oder Gestaltungsdateien als Bestandteile der Abbilder. Beispielhaft sind an dieser Stelle JavaScript-Frameworks externer Anbieter zu nennen.

Nachdem im vorliegenden Kapitel Module und Funktionen skizziert wurden, die das Erfassen von Abbildern ermöglichen, beschäftigt sich das nachfolgende Kapitel mit der Beschreibung von Modulen und Funktionen, die das Bewahren der Abbilder ermöglichen.

### 5.4.3 Bewahren

Zum Bewahren der Abbilder werden digitale Kopien der Web-Seiten in Datenbanken bzw. Dateispeicher geschrieben und mit Metadaten interpretierbar gemacht.<sup>402</sup>

Das Erschließen und Suchen der Abbilder auf Basis von deskriptiven, technischen und strukturellen Metadaten ist in diesem Zusammenhang unabdingbar.<sup>403</sup>

Einige Metadaten können während der Erfassung automatisiert definiert werden und andere müssen nachträglich bzw. manuell eingepflegt werden.<sup>404</sup> Beispielhaft sind an dieser

---

402 Vgl. Harvey, D. R.; Gillian, Oliver: Digital curation, a. a. O., S. 31 ff.

403 Vgl. Miller, Steven J.: Metadata for digital collections: a how-to-do-it manual, a. a. O., S. 26 ff.

404 Vgl. Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 39 f.

Stelle Metadaten über den organisatorisch-technischen Betrieb zu nennen. Diese Kontext-Informationen können nicht während der Erfassung aus Dateien, Quellcodes oder anderen technischen Komponenten abgeleitet werden – Identifizierungsmerkmale, Namen und Dateiformate dagegen schon.<sup>405</sup>

Abbildung 37 grenzt die Module zur Bewahrung im WSBS von den restlichen Modulen des WSBS ab.

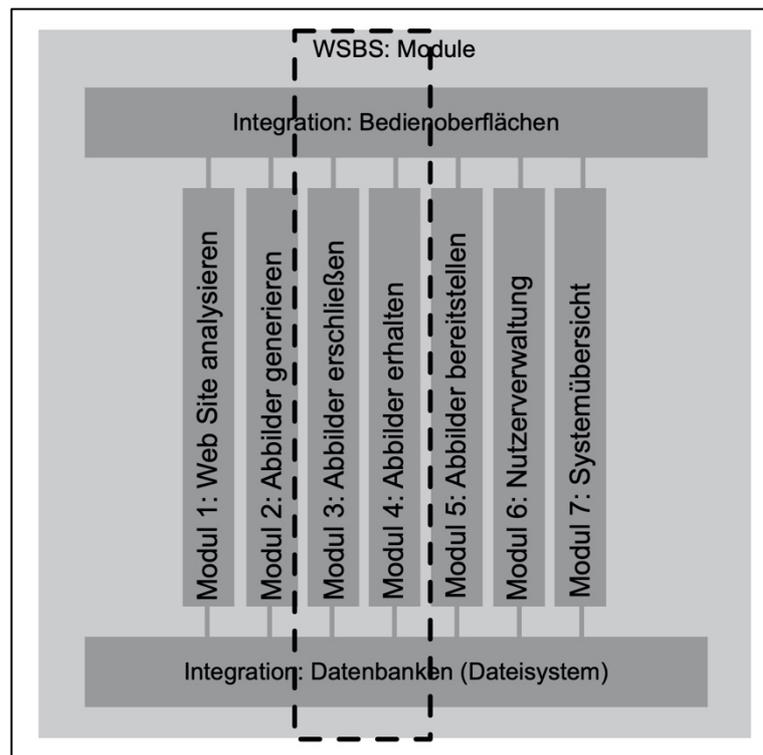


Abb. 37: Module zum Bewahren im WSBS

Die Module zum Bewahren gliedern sich in das „Abbilder erschließen“ und „Abbilder erhalten“. Mit Hilfe der Funktionen dieser Module soll eine Verwaltung der Abbilder ermöglicht werden.

Aus analytischen Gründen findet eine Untergliederung der Funktionen in Sub-Module statt. Zum „Abbilder erschließen“ werden die Sub-Module „Organisatorische Infrastruktur“ und „Technische Infrastruktur“ implementiert.

405 Vgl. Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.): Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, a. a. O., S. 106 f.

Abbildung 38 stellt die (Sub-)Module zum Erschließen der Abbilder in einem Funktionsbaum überblicksartig dar.

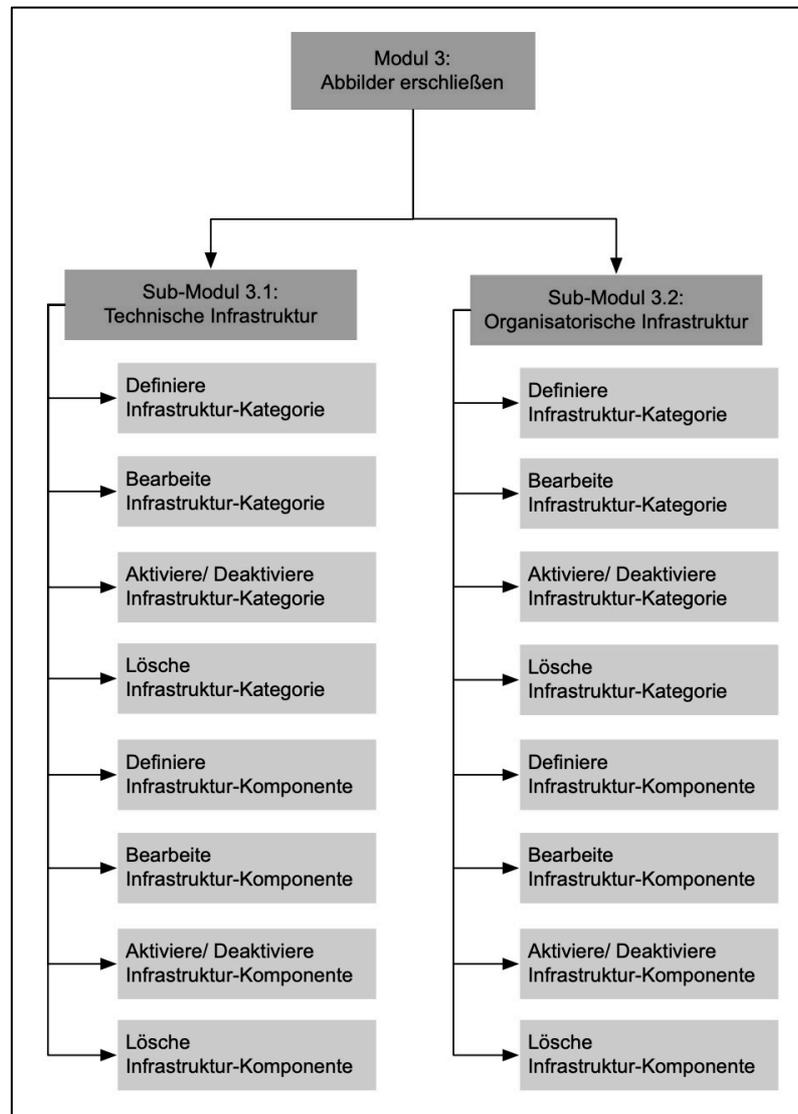


Abb. 38: Module und Funktionen zum Erschließen

Zum „Abbilder erschließen“ werden im Folgenden Sub-Module aufgelistet und Funktionen beschrieben:

- „Technische Infrastruktur“: Zur Dokumentation des betriebsspezifischen Kontexts werden technische Infrastruktur-Komponenten beschrieben. Die technischen Infrastruktur-Komponenten werden durch Kategorien strukturiert. Mit einer Kategorie wird festgelegt, um welche Komponente es sich handelt. Eine Infrastruktur-Komponente kann beispielsweise ein AWS, ein virtueller Server oder auch Server Hardware sein. Zu jeder Infrastruktur-Komponente wird ein Einsatz- bzw. Betriebszeitraum erfasst. Die Dokumentation der technischen Infrastruktur-

Komponente wird allen Abbildern zugeordnet, die im Einsatz- bzw. Betriebszeitraum generiert werden.

- „Organisatorische Infrastruktur“: Zur Dokumentation des betriebsspezifischen Kontexts werden organisatorische Infrastruktur-Komponenten beschrieben. Organisatorische Infrastruktur-Komponenten sind im vorliegenden Systementwurf Mitarbeiter, die zum Betrieb der Web Site eingesetzt werden. Ein Mitarbeiter kann z. B. ein Web-Redakteur, Systemadministrator oder IT-Verantwortlicher sein. Für jeden Mitarbeiter kann ein Tätigkeitszeitraum erfasst werden. Der Tätigkeitszeitraum umfasst die Zeitspanne, in der ein Mitarbeiter am Betrieb der Web Site mitgewirkt hat.

Typischerweise umfasst die Erschließung von Archivgut das Aufbereiten und Ergänzen fehlerhafter oder unvollständiger Informationen.<sup>406</sup> Da ein wesentlicher Teil der Metadaten bereits automatisiert durch Systemfunktionen erschlossen wird, befasst sich die manuelle Erschließung im vorliegenden Systementwurf zunächst mit den betriebsspezifischen Metadaten.

Neben dem Verwalten von betriebsspezifischen Metadaten stellt das WSBS Funktionen bereit, die Inhalt und Gestalt der Abbilder in der Speicherumgebung editierbar machen. Unter anderem, um Veränderungen vorzunehmen, die aus rechtlichen Aspekten resultieren, z. B. das Entfernen von sensiblen Mitarbeiterfotos.

Zum „Abbilder erhalten“ werden die Sub-Module „Abbilder-Editor“ und „Migrationsverwaltung“ implementiert.

---

<sup>406</sup> Vgl. auch hier Harvey, D. R.; Gillian, Oliver: Digital curation, a. a. O., S. 57 f. und Keitel, Christian; Deutsches Institut für Normung (Hrsg.): Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, a. a. O., S. 10 f.

Abbildung 39 stellt die (Sub-)Module und Funktionen zum Editieren der Abbilder in einem Funktionsbaum dar.

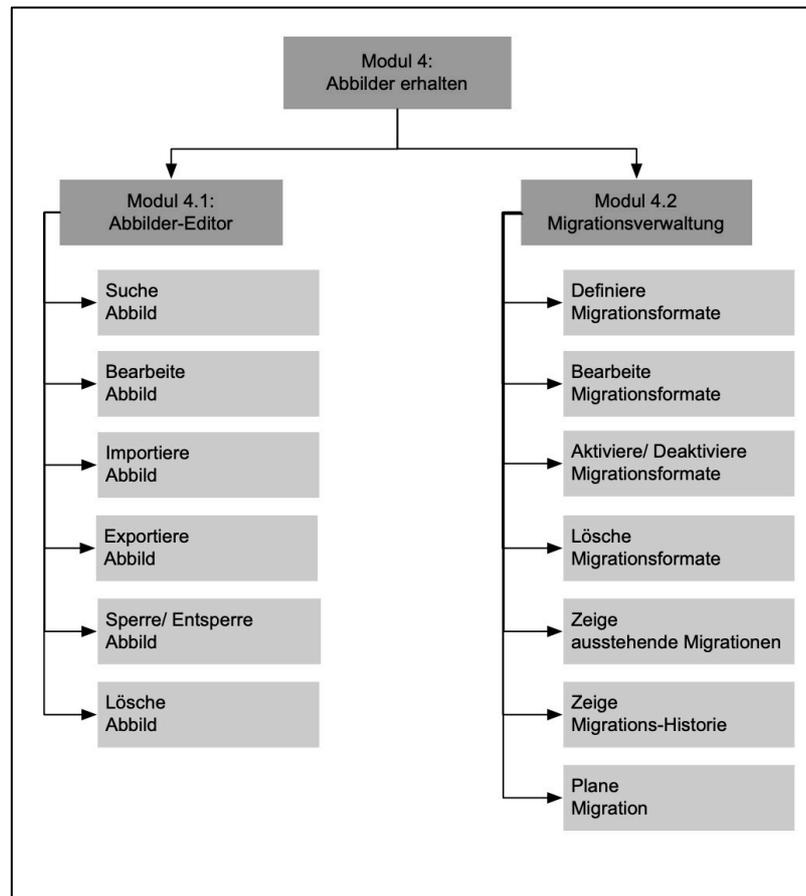


Abb. 39: Module und Funktionen zum Erhalten

Zum Modul „Abbilder erhalten“ werden im Folgenden die Sub-Module aufgelistet und Funktionen beschrieben:

- „Abbilder-Editor“: Das Erhalten von Inhalten bedingt unter Umständen, dass Änderungen an Abbildern vorgenommen werden müssen, um rechtlichen Vorgaben zu entsprechen. Einzelne Abbilder können daher nach der Speicherung durchsucht und bearbeitet werden. Das Bearbeiten der Abbilder umfasst Änderungen an den Inhalten eines Abbildes. Jede Änderung erzeugt eine neue Version des Abbildes und wird in den Metadaten dokumentiert. Authentizität und Integrität der Abbilder müssen stets gewahrt werden.<sup>407</sup> Darüber hinaus besteht die Möglichkeit, einzelne Abbilder zu sperren. Das Importieren eines individuellen Abbildes ist zudem möglich, um eine manuelle Erfassung von Zuständen der Web Site zu realisieren. Der Import muss die gleiche Struktur aufweisen, wie Abbilder, die

407 Vgl. Schumann, Natascha: Einführung in die digitale Langzeitarchivierung, a. a. O., S. 42.

automatisiert erfasst und bewahrt werden. Mit Hilfe einer Export-Funktion können Abbilder aus den Datenbanken kopiert und auf lokale Rechner oder andere IT-Systeme übertragen werden.

- „Migrationsverwaltung“: Die Migrationsverwaltung integriert eine Funktionsmenge, die das Durchführen von Migrationsaufgaben für einzelne Abbilder unterstützt. Migrationsformate können definiert, bearbeitet, (de-)aktiviert oder gelöscht werden. Auf Basis der Migrationsformate überprüft das System, ob Abbilder veraltete Dateiformate verwenden. Ausstehende Migrationen können geplant und durchgeführte Migrationen angezeigt werden.

Nachdem im vorliegenden Kapitel die Funktionen zur Bewahrung aufgezeigt wurden, widmet sich das folgende Kapitel der Bereitstellung von Abbildern.

#### 5.4.4 Bereitstellen

Das Bereitstellen im WSBS umfasst das Erlebarmachen der Abbilder einer Web Site im Web Browser eines Nutzers.

Abbildung 40 grenzt das Modul zur Bereitstellung von den restlichen Modulen ab.

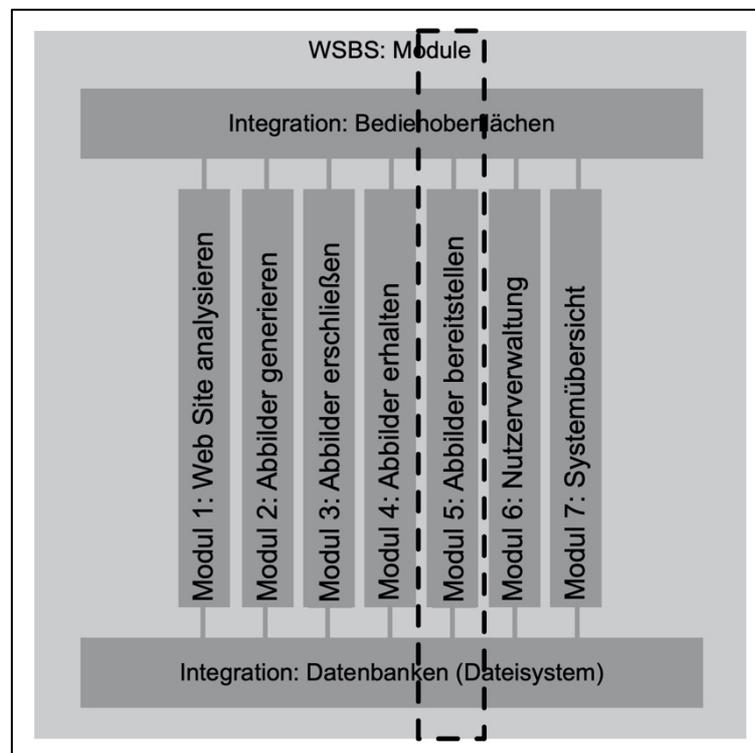


Abb. 40: Modul zum Bereitstellen im WSBS

Aus analytischen Gründen findet eine Untergliederung der Funktionen in Sub-Module statt. Zum „Abbilder bereitstellen“ werden die Sub-Module „Reproduktionsumgebung“ und „Kontextumgebung“ implementiert.

Abbildung 41 stellt die (Sub-)Module und Funktionen zum Bereitstellen der Abbilder einer WSB in einem Funktionsbaum dar.

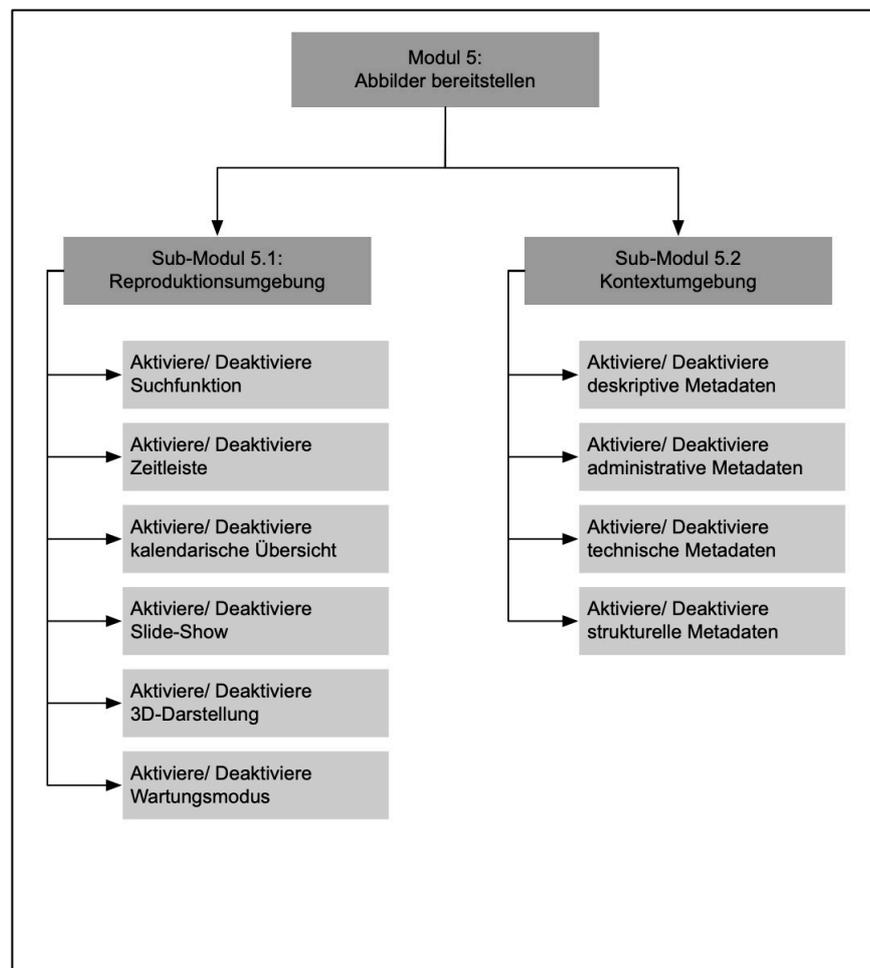


Abb. 41: Module und Funktionen zum Bereitstellen der WSB

Im Folgenden werden die Funktionen der Sub-Module zum Bereitstellen der Abbilder einer WSB aufgelistet und erläutert.

- „Reproduktionsumgebung“: Zum Erleben und Recherchieren der einzelnen Abbilder kann die Reproduktionsumgebung granular konfiguriert werden. Einzelne Elemente auf der Bedienoberfläche können zum Erlebbarmachen der WSB aktiviert oder deaktiviert werden. Mit Hilfe der Suchfunktion werden Abbilder aufgrund ihrer Metadaten auffindbar gemacht. Sowohl deskriptive, technische als auch strukturelle Metadaten können mit der Suche ausgewertet werden. Weiterhin ist die Navigation auf Basis einer Zeitleiste oder einer kalendarischen Übersicht

möglich. Zudem besteht die Option, unterschiedliche Darstellungsformen wie eine Slide-Show-basierte Navigation oder eine Drei-Dimensionale-Darstellung der Abbilder zu aktivieren.

- „Kontextumgebung“: Abbilder sind auf Basis ihrer Metadaten recherchierbar. Betreiber können durch das Aktivieren bzw. Deaktivieren deskriptiver, struktureller, administrativer und technischer Metadaten bestimmen, welche Informationen mit den Abbildern im Web veröffentlicht werden. Sensible Informationen können mit diesen Funktionen von der Veröffentlichung ausgeschlossen werden.

Die Metadaten werden im Informationsbereich der Nutzungsoberfläche angezeigt (vgl. dazu nachfolgende Abbildung 42).

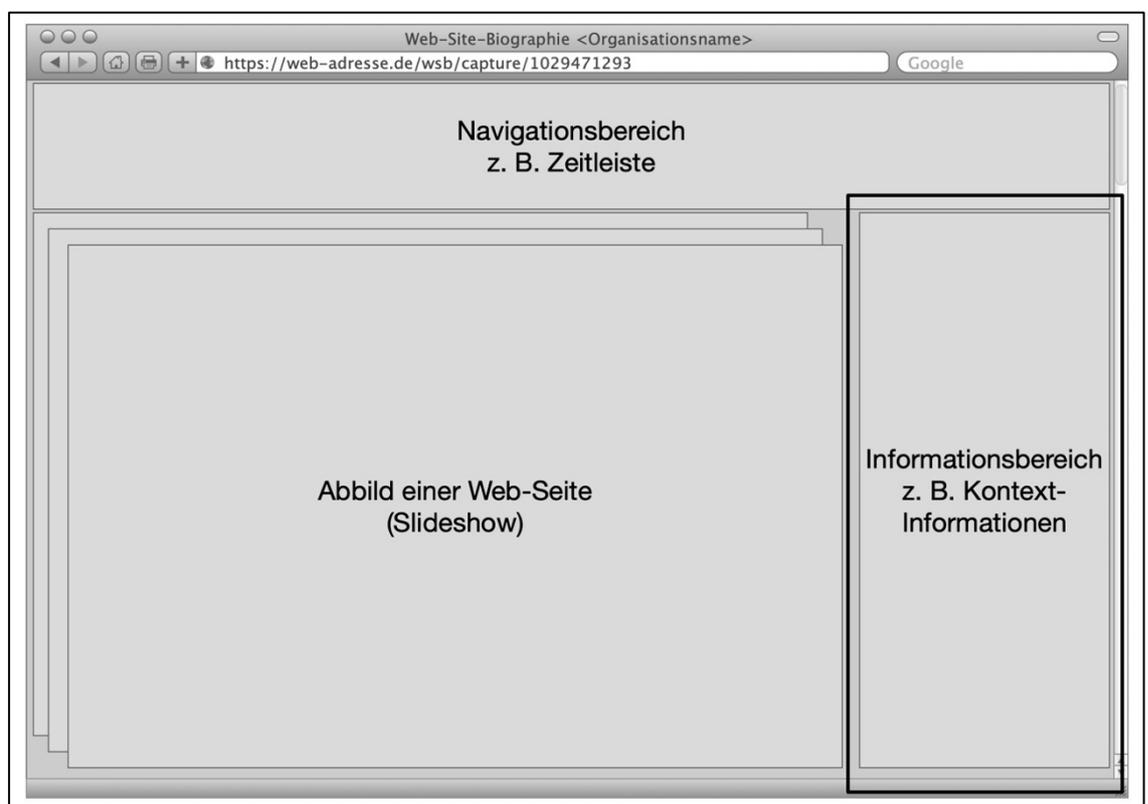


Abb. 42: Positionierung der Metadaten eines Abbildes im Web Browser

Aufbauend auf den Modulen und Funktionen zum Erfassen, Bewahren und Bereitstellen der Abbilder werden abschließend Funktionen zur Nutzerverwaltung und Systemübersicht skizziert und erläutert.

### 5.4.5 Nutzerverwaltung und Systemübersicht

Die Nutzerverwaltung übernimmt das Anlegen und Editieren von Benutzer-Konten. Zur Qualitätssicherung wird zudem eine Systemüberwachung integriert. Mit der Systemüberwachung werden Zusammenfassungen der Erfassungsprozesse und Fehlerberichte erstellt und so die Interpretationen der Systemaktivitäten ermöglicht.

Abbildung 43 grenzt die Nutzerverwaltung und Systemübersicht von den restlichen Modulen des WSBS ab.

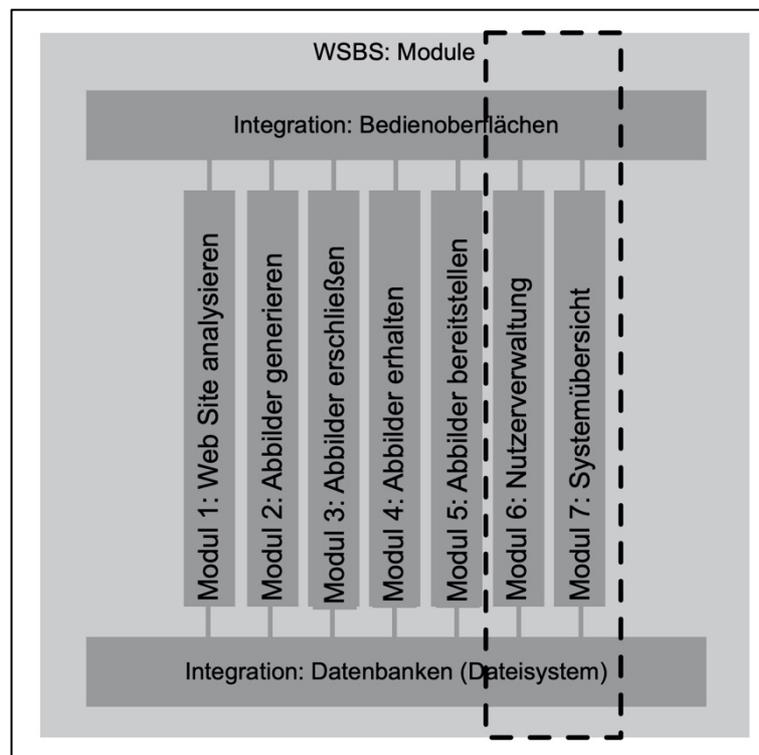


Abb. 43: Module zur Kontrolle des WSBS

Um eine übersichtliche Darstellung der Funktionen der Module „Nutzerverwaltung“ und „Systemübersicht“ zu erreichen, findet auf konzeptioneller Ebene eine Untergliederung in Sub-Module statt. Zur Nutzerverwaltung werden die Sub-Module „Benutzerkonfiguration“ und „Berechtigungsgruppen“ implementiert. Für die „Systemübersicht“ findet eine Untergliederung in die Sub-Module „Erfassung“ und „Bewahrung und Bereitstellung“ statt.

Abbildung 44 stellt die (Sub-)Module und Funktionen der „Nutzerverwaltung“ in einem Funktionsbaum dar.

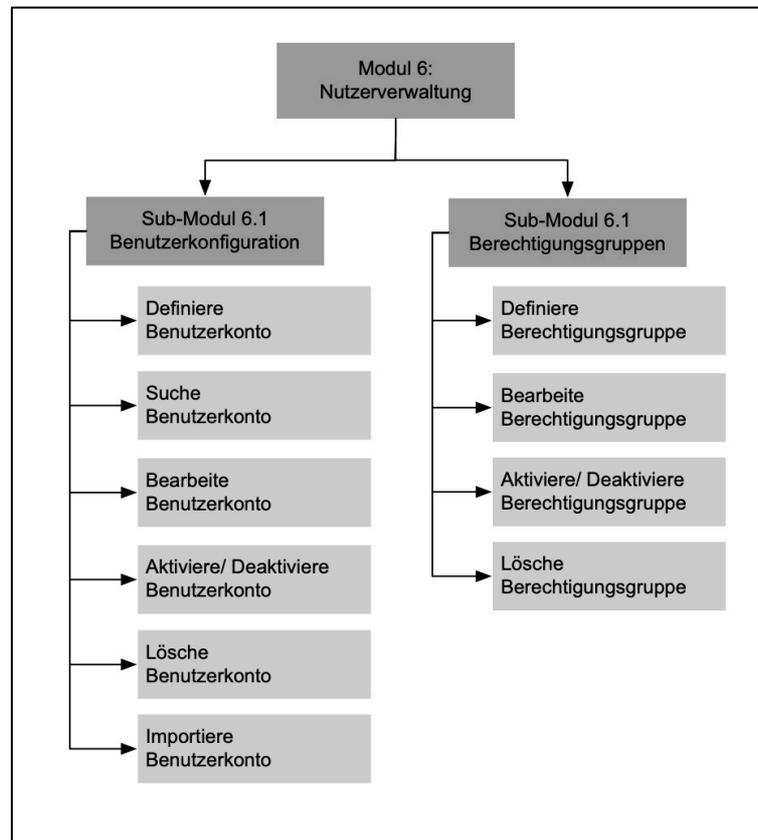


Abb. 44: Module und Funktionen zur Systemverwaltung (1/2)

Im Folgenden werden die Funktionen und (Sub-)Module zur Nutzerverwaltung aufgelistet und erläutert.

- „Benutzerkonfiguration“: Administratoren des WSBS können Benutzerkonten definieren, suchen, bearbeiten und löschen. Zudem besteht die Möglichkeit, Benutzerkonten über standardisierte Schnittstellen zu importieren. Wird das WSBS in eine IT-Infrastruktur integriert, die eine zentrale Benutzerkontenverwaltung implementiert, kann diese angebunden und verwendet werden.
- „Berechtigungsgruppen“: Da nicht nur innerhalb von Organisationen, sondern auch innerhalb von IT-Systemen divergierende Aufgabenbereiche und Verantwortlichkeiten existieren, werden Funktionen implementiert, die Berechtigungsgruppen erzeugen. Mit Berechtigungsgruppen können Handlungsspielräume einzelner Benutzerkonten z. B. auf Lese-, Schreib- oder Lösch-Rechte reduziert bzw. erweitert werden. Berechtigungen lassen sich pro Benutzerkonto definieren, bearbeiten, aktivieren, deaktivieren oder löschen.

Neben dem Verwalten von Benutzerkonten und Berechtigungen stehen Funktionen bereit, die Systemaktivitäten darstellen.

Abbildung 45 stellt die (Sub-)Module und Funktionen der „Systemübersicht“ in einem Funktionsbaum dar.

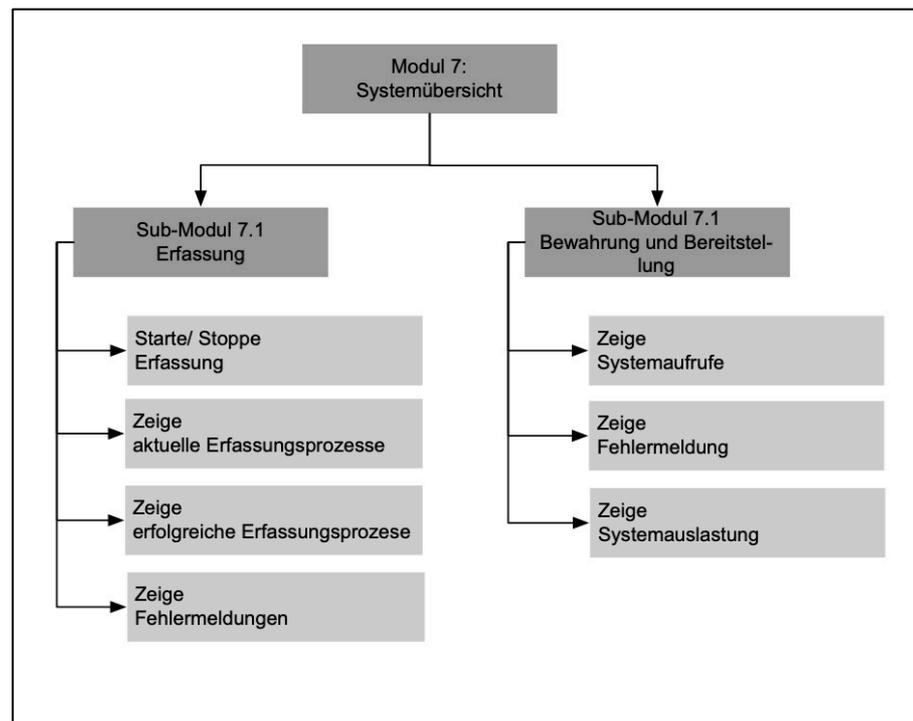


Abb. 45: Module und Funktionen zur Systemverwaltung (2/2)

Zum Modul „Systemübersicht“ werden im Folgenden die in Abbildung 45 aufgelisteten Sub-Module und Funktionen beschrieben:

- „Erfassung“: Die Systemübersicht der Erfassungsprozesse ermöglicht nicht nur das Starten sowie das Stoppen der Erfassungsprozesse, sondern stellt auch die Zustände der Erfassungsprozesse dar. Erfolgreiche Erfassungsprozesse werden kompakt aufbereitet und präsentiert. Fehlermeldungen, die z. B. in Log-Dateien dokumentiert werden, können analysiert werden.
- „Bewahrung und Bereitstellung“: Die Systemübersicht der Bewahrung und Bereitstellung ermöglicht das Überwachen von Systemaufrufen (z. B. Abfragen der Abbilder) und Fehlermeldungen. Die Auslastung des Gesamtsystems wird beispielsweise durch die Auslastung des zur Verfügung stehenden Speicherplatzes angezeigt.

Aufbauend auf der Funktionsbetrachtung folgt im abschließenden Kapitel eine Zusammenfassung und Erläuterung des Fortgangs der Arbeit.

## 5.5 Zusammenfassung und Erläuterung des Fortgangs der Arbeit

Im Rahmen des Systementwurfs wurde die Systemarchitektur auf Basis abgegrenzter Systembausteine entwickelt. Die Systembausteine übernehmen unterschiedliche Aufgaben. In Summe ermöglicht die Kombination aller Systembausteine das Erfassen, Bewahren und Bereitstellen der Lebensgeschichte einer Web Site (WSB).

Die Vollständigkeit des Systementwurfs wird durch das Adressieren und Skizzieren sämtlicher Komponenten des Strukturmodells belegt. Das Strukturmodell selbst wurde auf Basis etablierter Rahmenwerke abgeleitet und für das vorliegende Konstruktionsziel der Arbeit angepasst (vgl. Kapitel 2).<sup>408</sup>

Funktionen werden pro Modul konkretisiert. Aufgabe der *Systemverwaltung* ist es, die Funktionen der Module auf Bedienoberflächen bereitzustellen und nutzbar zu machen.

Abbildung 46 stellt das Strukturmodell des WSBS abschließend dar und hebt die Systemverwaltung durch einen Rahmen hervor.

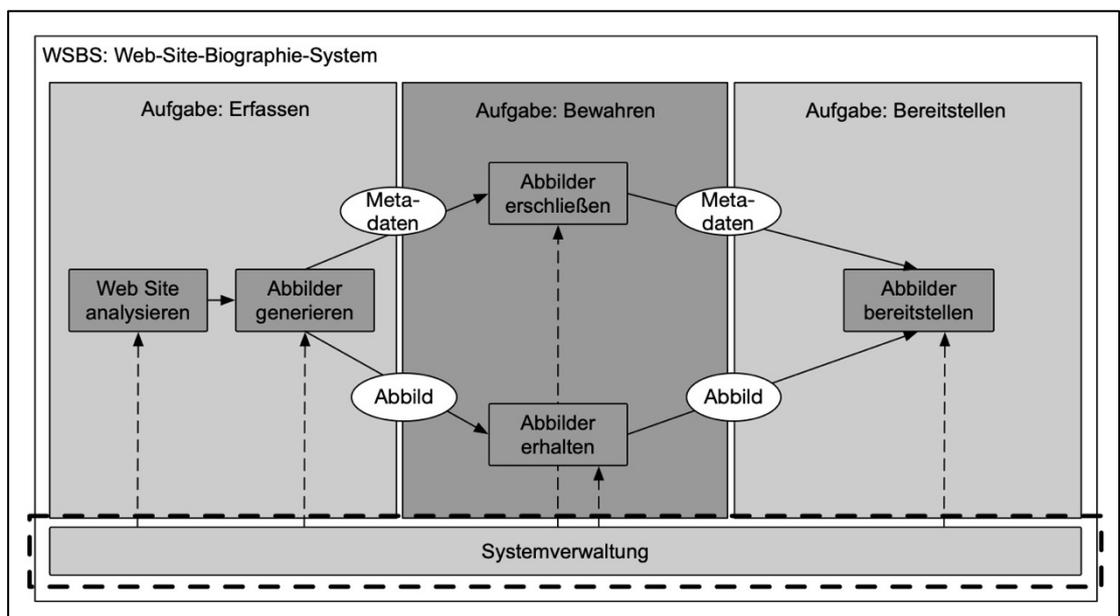


Abb. 46. Abschließende Darstellung des Strukturmodells

408 Vgl. Recker, Jan: Scientific research in information systems: a beginner's guide, a. a. O., S. 106 f.

## 6 Prototypische Realisierung eines WSBS

### 6.1 Systematisierung der prototypischen Realisierung

Aufbauend auf dem Systementwurf des Web-Site-Biographie-Systems (WSBS) wird im vorliegenden Kapitel die Realisierung des Prototyps beschrieben.<sup>409</sup> Es wird dabei in eine Beschreibung der Funktionsweise und Bedienoberflächen gegliedert.

Die Beschreibung der *Funktionsweise* ist durch Prozessschritte geprägt, die das Erfassen, Bewahren und Bereitstellen der WSB ermöglichen. Die einzelnen Prozessschritte werden in Kapitel 6.2 dargestellt und erläutert.

Im Anschluss an die Beschreibungen der Funktionsweise werden in Kapitel 6.3 *Bedienoberflächen* des WSBS visualisiert und dokumentiert. Durch das Visualisieren und Dokumentieren der Bedienoberflächen soll die Nutzung bzw. Bedienung des Systems nachvollziehbar gemacht werden.

### 6.2 Funktionsweise des WSBS

#### 6.2.1 Systematisierung der Funktionsweise des WSBS

Die Funktionsweise des WSBS ist in die Aufgaben *Erfassen*, *Bewahren*, *Bereitstellen* gegliedert. Im Prototyp des WSBS werden diese Aufgaben durch Anwendungskomponenten umgesetzt. Die Anwendungskomponenten wurden als Bestandteil der Systemarchitektur modelliert und lassen sich in eine Crawler-Anwendung und eine Curator-Anwendung gliedern.

Die Aufgabe der Crawler-Anwendung ist das Erfassen der WSB. Zentraler Bestandteil dieser Aufgabe ist das Analysieren von Veränderungen auf einer Web Site und das Generieren von Abbildern. Die Crawler-Anwendung wird im Folgenden auch als *WSBS-Crawler* bezeichnet.

Die Curator-Anwendung nimmt Abbilder und Metadaten des WSBS-Crawlers entgegen. Abbilder werden mit ihren Metadaten in einer Datenbank erschlossen und auf Basis von geeigneten Speichermedien für die Nachwelt konserviert und erlebbar gemacht. Die Curator-Anwendung wird im Folgenden auch als *WSBS-Curator* bezeichnet.

---

<sup>409</sup> Vgl. Pfeffers, Ken; Tuunanen, Tuure; Rothenberger, Marcus A.; Chatterjee, Samir: A design science research methodology for information systems research, a. a. O., S. 55 f.

Abbildung 47 hebt die Anwendungskomponenten in der schematischen Darstellung der Systemarchitektur mit einem Rahmen hervor.

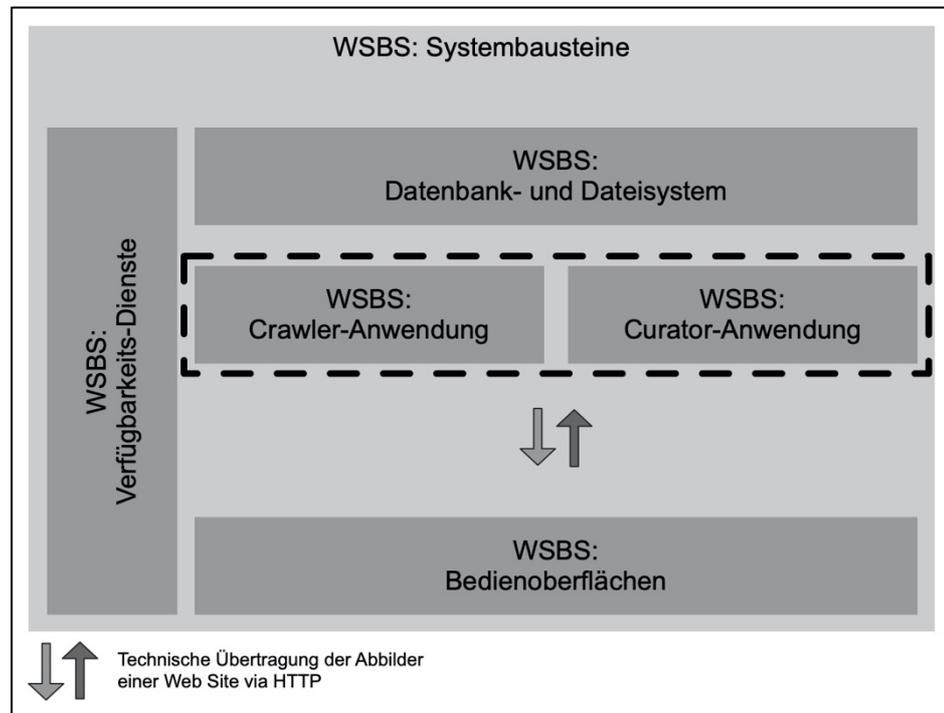


Abb. 47: Anwendungskomponenten zur Steuerung des WSBS

Die Funktionsweise des WSBS wird in den folgenden Kapiteln anhand des WSBS-Crawlers (Crawler-Anwendung) und WSBS-Curators (Curator-Anwendung) erläutert.

### 6.2.2 WSBS-Crawler

Das Erfassen der WSB wird durch das Analysieren der Web Site und das Generieren von Abbildern umgesetzt. Die Crawler-Anwendung des WSBS (kurz: WSBS-Crawler) muss die Web Site in geeigneten Abständen aufrufen und die einzelnen Web-Seiten der Web Site in Abbildern festhalten.

Das Generieren eines langfristig nutzbaren Abbildes setzt die Anwendung eines geeigneten Dateiformates voraus. Ein für die Langzeiterhaltung von Web Sites erprobtes Dateiformat ist z. B. das sogenannte „WARC-Format“.<sup>410</sup> Das WARC-Format wird in den unterschiedlichsten Archivierungs-Initiativen eingesetzt und eignet sich auf Grund seiner

<sup>410</sup> Vgl. Zierau, Eld; Peyrard, Sébastien: Digital Preservation Metadata in a Metada Ecosystem, in: Digital preservation metadata for practitioners, New York, NY: Springer Berlin Heidelberg 2016, S. 200 f. und International Internet Preservation Consortium: The WARC Format 1.0, Online im Internet: <https://iipc.github.io/warc-specifications/specifications/warc-format/warc-1.0/#file-and-record-model>, 25.01.2020.

ISO-Standardisierung für die Erhaltung von Abbildern einer Web Site.<sup>411</sup> Neben dem Einsatz des WARC-Formats sind aber auch weitere Dateiformate für die Erhaltung von digitalen Kopien einer Web Site geeignet, z. B. Dateien im HTML- oder XML-Format.<sup>412</sup> Für den vorliegenden Prototyp werden Abbilder einzelner Web-Seiten zunächst in ihrem ursprünglichen Format – dem HTML-Format – aufbewahrt. Durch das Beibehalten des HTML-Formats wird eine Konvertierung in ein spezielles Format ausgespart und so ein Performance-Gewinn in der Verarbeitung möglich. Eine Konvertierung in weitere Dateiformate ist jederzeit umsetzbar.

Neben der Implementierung eines geeigneten Dateiformates ist das Implementieren von Mechanismen notwendig, die Zustandsänderungen einer Web Site verlässlich erkennen und in *vollständigen* Abbildern festhalten. Der WSBS-Crawler muss daher sämtliche Inhalte und Gestaltungselemente einer jeden Web-Seite möglichst unmittelbar nach einer Zustandsänderung kopieren. Problematisch ist dabei unter anderem der allgegenwärtige Einsatz von JavaScript (kurz: JS) und das Nachladen von Inhalten, wie Navigationsmenüs, Bildern oder Texten.

Einfache Crawler-Anwendungen öffnen im Regelfall automatisiert alle Web-Seiten einer Web Site und erstellen Abbilder zum Zeitpunkt des Aufrufes. Werden Inhalte oder Gestaltungselemente durch den Einsatz von JS wenige Sekunden nach dem Aufruf oder durch Nutzerinteraktion nachgeladen, sind diese kein Bestandteil des Abbildes.<sup>413</sup> Da heute aber eine Vielzahl von Web-Seiten durch dynamische und nachladbare Elemente realisiert werden, resultieren die Abbilder klassischer Crawler nicht immer in vollständigen Kopien einzelner Web-Seiten: Funktionen oder Gestaltungselemente fehlen. Für den Prototyp des WSBS wurde daher eine Kombination aus einem Crawler-basierten und transaktionalen Ansatz gewählt. Dieser *hybride* Ansatz soll das Problem unvollständiger Abbilder lösen und sowohl Nutzerinteraktionen als auch das Nachladen von Bildern in einem Abbild berücksichtigen.<sup>414</sup>

Damit das WSBS eine vollständige Erfassung der Abbilder einer Web Site realisieren kann, vereint der WSBS-Crawler die Erfassungstechniken eines klassischen Web Crawlers mit den Techniken der transaktionalen Erfassung. Auf Basis dieses Techniken-Mix ist es dem Prototyp möglich, alle Inhalte und Gestaltungselemente einer Web-Seite zu

---

411 Vgl. Zierau, Eld; Peyrard, Sébastien: Digital Preservation Metadata in a Metada Ecosystem, a. a. O., S. 200 f. und International Internet Preservation Consortium: The WARC Format 1.0, a. a. O., 25.01.2020.

412 Vgl. Ziehl, Stefan: Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, a. a. O., S. 58 ff.

413 Vgl. z. B. Masanès, Julien: Web Archiving: Issues and Methods, a. a. O., S. 26.

414 Vgl. Menczer, Filippo: Web Crawling, a. a. O., S. 274 f.

kopieren – auch solche Inhalte und Gestaltungselemente, die erst nach einem bestimmten Zeitraum oder durch Nutzerinteraktion nachgeladen werden.

Abbildung 48 stellt die Funktionsweise der Erfassung in zwei Schritten dar.

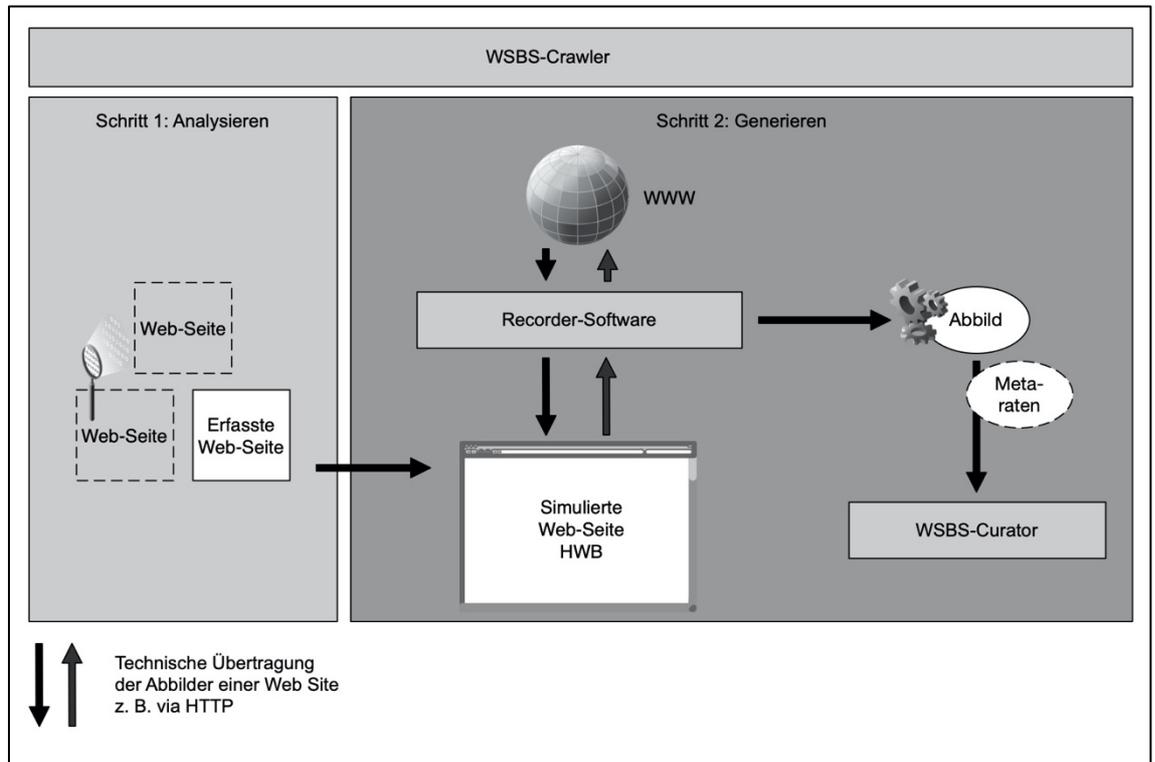


Abb. 48: Erfassungsprozess und Komponenten des WSBS-Crawlers

Die Erfassung mit dem WSBS-Crawler wird im Folgenden erläutert.

- Schritt 1 „Analysieren“: Damit der WSBS-Crawler vollständige Abbilder einer Web Site erzeugen kann, müssen alle Web-Seiten im Adressraum der Organisation erfasst werden. Ausgehend von einer Startseite öffnet der WSBS-Crawler alle im WWW veröffentlichten Web-Seiten, die dem Adressraum der Organisation zugeordnet sind. Der WSBS-Crawler bewegt sich dabei von einer Web-Seite zur nächsten. Das System erfasst die Links im HTML-Quellcode und wandert entlang dieser virtuellen Pfade, um die Web Site einer Organisation möglichst vollständig zu indexieren.<sup>415</sup> Jede Web-Seite wird geöffnet und auf Verlinkungen zu weiteren Web-Seiten überprüft. Am Ende des Erfassungsvorgangs entsteht ein Index, der Web-Adressen der zu kopierenden Web-Seiten enthält.

<sup>415</sup> Vgl. Pant, Gautam; Srinivasan, Padmini; Menczer, Filippo: *Crawling the Web*, a. a. O., S. 153. oder Alpar, Andre; Koczy, Markus; Metzen, Maik: *SEO - Strategie, Taktik und Technik: Online-Marketing mittels effektiver Suchmaschinenoptimierung*, a. a. O., S. 47 f.

- Schritt 2 „Generieren“: Zum Generieren eines Abbildes wird eine Web-Adresse aus dem Index geladen und die original Web-Seite geöffnet. Der WSBS-Crawler verwendet zum Öffnen der Web-Seite einen „Headless Web Browser (kurz: HWB)“.<sup>416</sup> Das Besondere an dieser Browser-Software ist, dass sie als Bestandteil einer Server-Anwendung verwendet werden kann und auf die Darstellung einer Web-Seite in einem Browser-Fenster verzichtet. Die Interpretation, Nutzung und Bearbeitung von Web-Seiten bzw. HTML-Dateien erfolgt jedoch im gleichen Maße, wie bei einem regulären Web Browser. Es steht lediglich kein Fenster, sondern eine Kommandozeile zum Bedienen zur Verfügung. Über die Kommandozeile oder geeignete Programmierschnittstellen können einzelne Web-Seiten geladen, Links geklickt oder auch Bildschirmfotos von einer Web-Seite erzeugt werden. Der HWB ermöglicht damit nicht nur das einfache Öffnen einer Web-Seite, sondern stellt dem Prototyp Funktionen bereit, die das Simulieren von Nutzerverhalten ermöglichen. Mit Programm-Code werden dem HWB Befehle zum Öffnen einer Datei, zum Abspielen einer Slide-Show, zum Öffnen eines dynamischen Menüs oder zum Scrollen mitgegeben. Auf diese Weise lassen sich Inhalte, die klassische Web Crawler nicht abbilden können, anzeigen und sichern. Das Sichern sämtlicher Inhalte und Gestaltungselemente setzt eine *Recorder-Software* voraus. Während eine Web-Seite im HWB geladen und angezeigt wird, werden alle Inhalte und Gestaltungselemente aufgezeichnet. Der WSBS-Crawler spiegelt jede Transaktion, die durch den HWB ausgelöst wird. Die Recorder Software zeichnet solange alle übertragenen Elemente (Skripte, Bilder, Texte etc.) auf, bis eine Web-Seite durch den WSBS-Crawler wieder geschlossen wird. Auf Basis der übertragenen Elemente kann eine Web-Seite vollständig kopiert und zum Erleben der Vergangenheit reproduziert werden. Nachdem sämtliche Transaktionen durch die Recorder-Software aufgezeichnet wurden, werden die kopierten Elemente als Abbild an den WSBS-Curator übergeben. Der WSBS-Curator beinhaltet die Bewahrungs- und Bereitstellungskomponenten und wird im folgenden Kapitel erläutert.

---

<sup>416</sup> Vgl. zum Begriff und Einsatz eines „Headless Web Browsers“ Mozilla and individual contributors (Hrsg.): Headless mode, Online im Internet: [https://developer.mozilla.org/en-US/docs/Mozilla/Firefox/Headless\\_mode](https://developer.mozilla.org/en-US/docs/Mozilla/Firefox/Headless_mode), 25.01.2020.

### 6.2.3 WSBS-Curator

Die Aufgabe des WSBS-Curators ist das Bewahren und Bereitstellen der Abbilder. Die Abbilder einzelner Web-Seiten werden vom WSBS-Crawler an den WSBS-Curator übergeben, für die weitere Verwendung erschlossen und anschließend bereitgestellt.

Abbildung 49 stellt die Schritte zum Erschließen und Bereitstellen der Abbilder mit dem WSBS-Curator dar. Die Nummerierung der Schritte schließt an die Schritte des Erfassungsprozesses an.

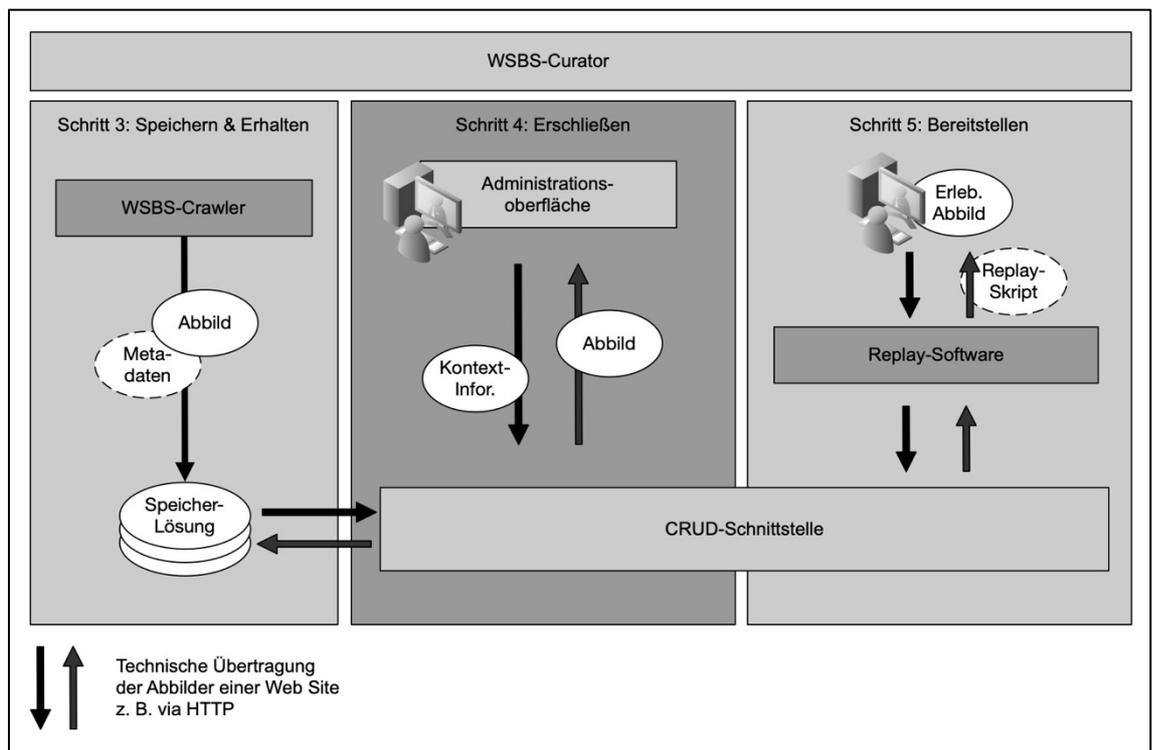


Abb. 49: Bewahren der Abbilder mit dem WSBS-Curator

- Schritt 3 „Speichern und Erhalten“: Die Abbilder einer Web Site werden durch die Recorder-Software in einem standardisierten Datei-Format an den WSBS-Curator übergeben. Das Abbild einer Web-Seite wird anschließend einer Sammlung von Abbildern in der Speicherlösung hinzugefügt. Der Speicherort des neu hinzugefügten Abbildes wird ausgelesen und in einem „Abbilder-Index“ festgehalten. Der Abbilder-Index ist ein Bestandteil der Speicherlösung und kann als Register verstanden werden, in dem die Speicherorte aller Abbilder des WSBS notiert werden. Mit Hilfe dieses Registers kann das WSBS einzelne Abbilder

finden und zur Reproduktion aus der Sammlung laden. Das Erzeugen und Verwalten der Sammlung sowie des Registers geschieht automatisiert und wird mit einer *CouchDB*-Installation<sup>417</sup> implementiert.

- Schritt 4 „Erschließen“: Nach der Aufnahme eines Abbildes in die Speicherlösung können Kontext-Informationen zur Beschreibung des betriebsspezifischen Umfeldes durch einen Administrator des WSBS eingepflegt werden. Das betriebsspezifische Umfeld umfasst Beschreibungen der technischen und organisatorischen IT-Infrastruktur der Web Site. Einzelne Infrastruktur-Komponenten wie Server-Hardware und Software können erfasst und z. B. mit Angaben zum Einsatzzeitraum gespeichert werden. Diese Kontext-Informationen werden automatisch allen Abbildern zugeordnet, die im entsprechenden Einsatzzeitraum erfasst wurden. Das Erstellen der Kontext-Informationen erfolgt mit Formularen, die auf einer Bedienoberfläche zur Administration bereitgestellt werden. Die Bedienoberfläche ist über eine CRUD-Schnittstelle (Create, Read, Update, Delete; kurz: CRUD) mit der Speicherlösung verbunden. Diese Schnittstelle ermöglicht alle notwendigen Operationen, um das Erhalten und Auslesen von Abbildern und Metadaten umzusetzen. Die Bedienoberfläche kann losgelöst von der Speicherlösung entwickelt und über die Schnittstelle an den Bestand der Abbilder angebunden werden. Es wird so eine strikte Trennung zwischen der Präsentations- und Datenhaltungs-/Anwendungsschicht erreicht.

Nachdem die Schritte zum Speichern und Erschließen der Abbilder im WSBS erläutert wurden, befasst sich der nachstehende Abschnitt mit der Bereitstellung.

---

417 Vgl. The Apache Software Foundation (Hrsg.): CouchDB, Online im Internet: <https://couchdb.apache.org/>, 13.05.2020.

Abbildung 50 stellt den Schritt zum Bereitstellen der Abbilder mit dem WSBS-Curator dar. Die Nummerierung schließt an den Schritt der Erschließung an.

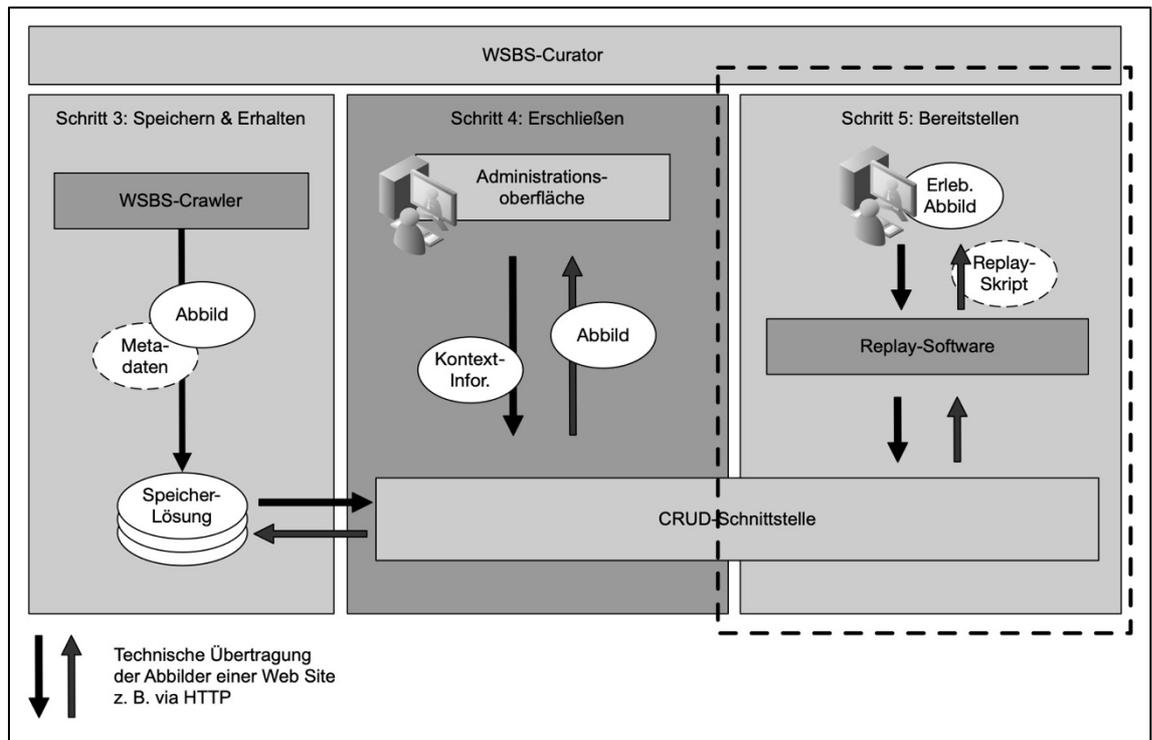


Abb. 50: Bereitstellen der Abbilder mit dem WSBS-Curator

- Schritt 5 „Bereitstellen“: Nachdem ein Abbild gespeichert und erschlossen wurde, kann es bereitgestellt werden. Die Bereitstellung eines Abbildes wird durch die „Replay-Software“ des WSBS-Curators möglich. Die Replay-Software sucht mit Hilfe der CRUD-Schnittstelle in der Speicherlösung nach Abbildern, die zu einer Nutzeranfrage passen. Wird ein passendes Abbild in der Speicherlösung identifiziert, lädt die Replay-Software zunächst eine HTML-Datei aus dem Speicher. Die HTML-Datei wird auf Basis des HTTP übertragen und im Web Browser des Nutzers zu einem vollständigen Abbild zusammengefügt. Zentraler Bestandteil des Zusammenfügens ist ein Replay-Skript<sup>418</sup>, welches von der Replay-Software übertragen wird. Das Replay-Skript wird automatisch im Web Browser des Nutzers installiert und dort dauerhaft ausgeführt. Es identifiziert alle Transaktionen, die der Web Browser auf Basis der HTML-Datei initiiert und leitet diese an das WSBS weiter. Style-Sheets, JS, Bilder oder Animationen, die erst nach Nutzerinteraktionen angefragt werden, lassen sich durch das

418 Das Replay-Skript wird als Service Worker implementiert. Vgl. hierzu z. B. Amarasinghe, Sean: Service worker development cookbook: build highly available and performant native web applications that seamlessly integrate with third-party APIs, 2016, S. 2 ff.

Replay-Skript identifizieren und passende Kopien aus der Speicherumgebung des WSBS laden. Durch das dauerhafte Analysieren der Transaktionen kann eine vollständige Reproduktion einzelner Web-Seiten gewährleistet werden. Im Gegensatz zu Web-Archiven, die den Quellcode eines Abbildes nur einmal während der Speicherung oder Auslieferung mit Verlinkungen zu internen Archiv-Ressourcen modifizieren, kann durch das Replay-Skript auch nachgeladener Inhalt ausgeliefert werden. Das erlebbare Abbild setzt sich also aus der HTML-Datei und allen durch das Replay-Skript nachgeladenen Inhalten und Gestaltungselementen zusammen.

Aufbauend auf der Beschreibung des WSBS-Curators soll im nächsten Kapitel eine Erläuterung einzelner Funktionen erfolgen. Die Erläuterung der Funktionen wird durch die Visualisierung und Beschreibung ausgewählter Bedienoberflächen durchgeführt.

## 6.3 Bedienoberflächen

### 6.3.1 Systematisierung der Bedienoberflächen

Die Bedienoberflächen machen Funktionen nutzbar, die zur Erstellung, Pflege und zum Erleben einer WSB benötigt werden. Die Menge der Bedienoberflächen leitet sich aus den Modulen des WSBS ab. Die Module selbst sind in das „Web Site analysieren“, „Abbilder generieren“, „Abbilder erschließen“, „Abbilder erhalten“ und „Abbilder bereitstellen“ gegliedert. Zudem sind im Systementwurf Module zur „Nutzerverwaltung“ und „Systemübersicht“ enthalten.

Abbildung 51 visualisiert die Module aus dem Systementwurf des WSBS.

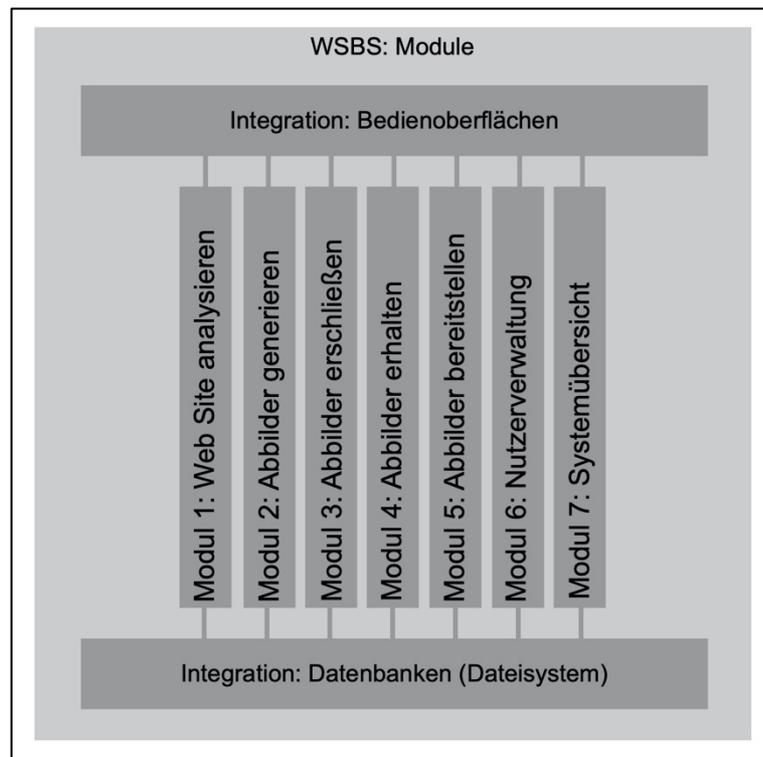


Abb. 51: Module des WSBS

Zum Bedienen der in Abbildung 51 skizzierten Module werden grafische Oberflächen benötigt, die die Funktionen der Module nutzbar machen.

Damit eine WSB erfasst werden kann, müssen die Funktionen der Module „Web Site analysieren“ und „Abbilder generieren“ in einer Bedienoberfläche eingebettet werden. Für das *Bewahren* einer WSB werden Bedienoberflächen benötigt, die die Funktionen der Module „Abbilder erschließen“ und „Abbilder erhalten“ umfassen. Ergänzend zu den Bedienoberflächen der Erfassung und Bewahrung sind Bedienoberflächen für die Module „Abbilder bereitstellen“, „Nutzerverwaltung“ und „Systemübersicht“ notwendig. Eine detaillierte Beschreibung dieser Bedienoberflächen erfolgt in den Kapiteln 6.3.2 bis 6.3.5.

Da sich der Prototyp des WSBS auf das *originalgetreue Reproduzieren* einer organisationalen Web Site fokussiert, werden in den folgenden Kapiteln nicht sämtliche Bedienoberflächen des WSBS aufgezeigt. Es werden vor allem Funktionen bzw. Bedienoberflächen gezeigt, die zum originalgetreuen Reproduzieren wesentlich sind.

### 6.3.2 Bedienoberflächen zum Erfassen einer WSB

Das Erfassen im WSBS wird durch Bedienoberflächen zum Analysieren der Web Site und zum Generieren von Abbildern ermöglicht. Zum Analysieren einer Web Site muss zunächst ein Adressraum auf einer Bedienoberfläche definiert werden. Der Adressraum bestimmt, unter welchen Web-Adressen die einzelnen Web-Seiten einer organisationalen Web Site abgerufen und kopiert werden.

Der Adressraum wird im WSBS durch folgende Angaben definiert:

- Startseite
- Prüfwert
- Web-Site-Betreiber
- Beschreibung

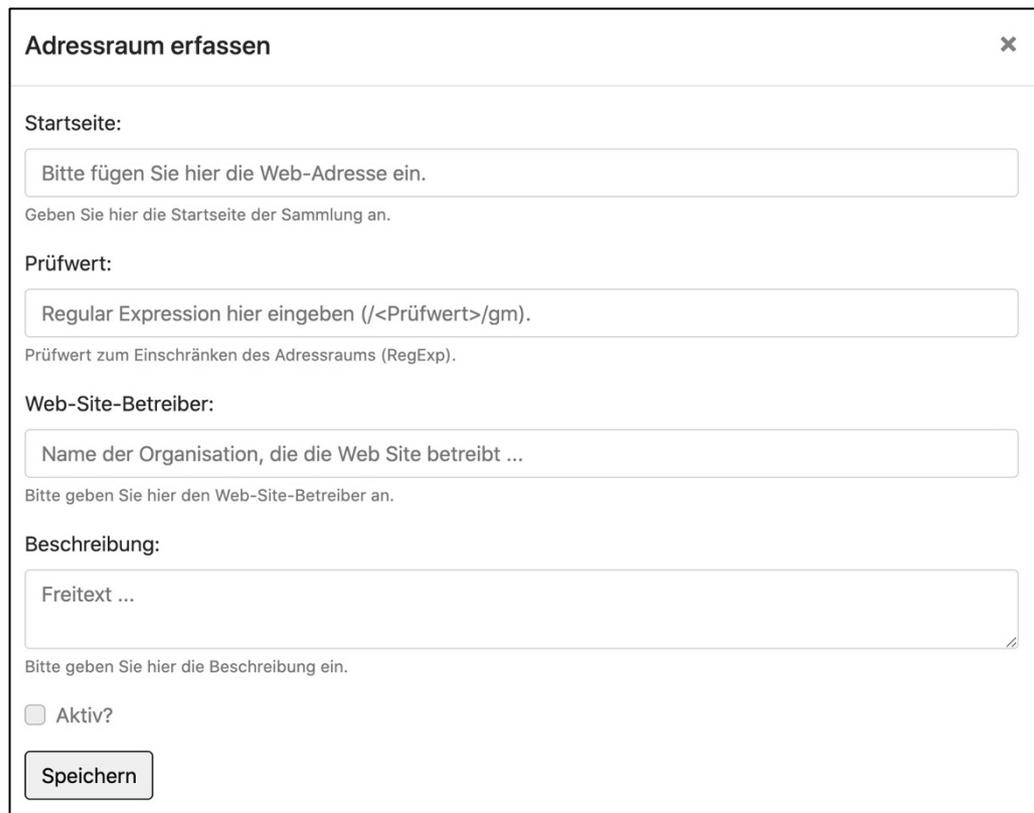
Die *Startseite* bestimmt, auf welcher Web-Seite das WSBS den Sammel- und Kopiervorgang startet. Ausgehend von dieser Startseite werden alle weiteren Web-Seiten der Organisation identifiziert und kopiert.

Mit einem *Prüfwert* kann festgelegt werden, ob alle Web-Seiten der Web Site kopiert werden sollen oder, ob die Web-Seiten eine bestimmte Zeichenkombinationen (Prüfwert) in ihrer Web-Adresse aufweisen müssen, damit sie als Bestandteil der WSB kopiert werden. Nur Web-Adressen, die den Prüfwert beinhalten, werden durch das WSBS verarbeitet.<sup>419</sup> Wird kein Prüfwert angegeben, sammelt das System alle Web-Seiten im Adressraum der Organisation. Das Verwenden eines Prüfwertes ist unter anderem dann notwendig, wenn das WSBS Abbilder einer Web Site erzeugen soll, die Teil einer dezentral organisierten Web-Präsenz sind. Beispielhaft können hier Web Sites einzelner Professuren innerhalb der Web-Präsenz einer Universität genannt werden. Alle Web Sites einer Universität werden im Regelfall unter der Domain der Universität aufgerufen. Damit das WSBS nicht alle Web Sites aller Organisationseinheiten der Universität sammelt, kann durch den Prüfwert festgelegt werden, dass nur Web-Seiten berücksichtigt werden, die eine bestimmte Zeichenkombination in ihrer Web-Adresse aufweisen, z. B. den Namen einer Professur. Auf diese Weise kann eine *organisationszentrierte* WSB erzeugt und nur Web-Seiten einer festgelegten Professur gesammelt werden. Ergänzend zur Angabe eines Prüfwertes muss ein Betreiber und eine Beschreibung definiert werden.

---

<sup>419</sup> Der Prüfwert wird als Regular Expression angegebene. Vgl. dazu z. B. Stubblebine, Tony: Regular expression pocket reference, 2. Ausgabe, Sebastopol, Calif. O'Reilly 2007, S. 3 ff.

Abbildung 52 stellt die Eingabemaske zum Erfassen eines Adressraums dar.



The image shows a web form titled "Adressraum erfassen" with a close button (X) in the top right corner. The form contains the following fields and labels:

- Startseite:** A text input field with the placeholder text "Bitte fügen Sie hier die Web-Adresse ein." Below it, a smaller text label reads "Geben Sie hier die Startseite der Sammlung an."
- Prüfwert:** A text input field with the placeholder text "Regular Expression hier eingeben (/<Prüfwert>/gm)." Below it, a smaller text label reads "Prüfwert zum Einschränken des Adressraums (RegExp)."
- Web-Site-Betreiber:** A text input field with the placeholder text "Name der Organisation, die die Web Site betreibt ...". Below it, a smaller text label reads "Bitte geben Sie hier den Web-Site-Betreiber an."
- Beschreibung:** A larger text input field with the placeholder text "Freitext ...". Below it, a smaller text label reads "Bitte geben Sie hier die Beschreibung ein."

At the bottom of the form, there is a checkbox labeled "Aktiv?" and a button labeled "Speichern".

Abb. 52: Eingabemaske zum Erfassen eines Adressraums

Da sich der Adressraum einer Web Site z. B. nach einem Relaunch der IT-Infrastruktur oder einer Unternehmensübernahme ändern kann, besteht die Möglichkeit, eine beliebige Anzahl an Adressräumen hinzuzufügen und zu verwalten.

Auf der folgenden Seite wird die Bedienoberfläche zum Verwalten der Adressräume im Prototyp des WSBS („wesby“) dargestellt.

Abbildung 53 stellt die Übersichtsseite zum Darstellen der Adressräume dar.

wesby pre-alpha v0.8

wesby Abbilder erfassen ▾ Abbilder bewahren ▾ Abbilder bereitstellen ▾ Systemübersicht ▾

## Adressraumverwaltung

Erstellen Sie Adressräume und erzeugen Sie Kopien einer Web Site.  
Gesammelte Web-Seiten im System: 35.752

[Adressraum hinzufügen](#)

[zu den Aktivitäten](#)

Löschen deaktiviert

Web Site	Gesammelte Web-Seiten	Zuletzt gesammelt
<p><b>https://www.uni-giessen.de/fbz/fb02</b></p> <p>Adressraum: <i>www.uni-giessen.de</i></p> <p>Prüfwert: fb02</p> <p>Web-Site-Betreiber: Fachbereich 02, JLU-Gießen</p> <p>Beschreibung: <i>Web Site des Fachbereichs 02 der Justus-Liebig-Universität Gießen.</i></p>	35.752 Web-Seiten	21/06/2020, 10:01:56 Uhr

[Bearbeiten](#)

[Jetzt sammeln](#)

Abb. 53: Adressräume im WSBS

Auf der Übersichtsseite in Abbildung 53 wird der Adressraum der Web Site des Fachbereichs 02 der JLU-Gießen dargestellt.

Ausgangspunkt für das Erfassen der Abbilder der Web Site ist die Startseite *https://www.uni-giessen.de/fbz/fb02*.

Nach der Anlage des Adressraums kann die Erfassung gestartet werden. Durch Klick auf den Button „Jetzt sammeln“ wird zunächst eine manuelle Sammlung angestoßen.

Abbildung 54 stellt das Fenster zum Starten einer manuellen Sammlung dar.

Sammlung von <https://www.uni-giessen.de/fbz/fb02> starten

Adressraum

Startseite: **https://www.uni-giessen.de/fbz/fb02**

Prüfwert: fb02

Parameter der Sammlung

Timeout in ms:

Auflösung in px:

Max. Dateigröße in MB:

Anzahl Web-Seiten:

Abb. 54: Manuelle Sammlung von Web-Seiten im WSBS

Die manuelle Sammlung einer Web Site muss durch Sammlungs-Parameter spezifiziert werden. Die Parameter der Sammlung sind ein Timeout, eine maximale Dateigröße in Megabyte (MB) und die Anzahl der Web-Seiten, die gesammelt werden sollen.

Mit einem *Timeout* wird festgelegt, wie lange das WSBS auf die Auslieferung einzelner Inhalts- und Gestaltungselemente einer Web Site wartet, bevor diese von der Erfassung ausgeschlossen werden. Eine Festlegung des Timeouts ist zwingend notwendig, damit das WSBS nicht unbegrenzt auf die Auslieferung einer einzelnen Ressource wartet und dadurch der Erfassungsprozess nicht abgeschlossen werden kann.

Eine Angabe der *maximalen Dateigröße in MB* legt fest, wie groß die Dateien einzelner Inhalte oder Gestaltungselemente sein müssen, damit sie als Bestandteil der WSB aufgenommen werden. Die Angabe der maximalen Dateigröße ist wesentlich für die System-Performance. Durch die Dateigröße kann verhindert werden, dass das WSBS Dateien kopiert, deren Kopiervorgang auf Grund der Dateigröße nicht abgeschlossen werden kann oder durch eine Zeitüberschreitung abgebrochen werden muss.

Die Anzahl der Web-Seiten pro Sammelvorgang ist abschließend festzulegen, damit das WSBS nur eine ausgewählte Menge an Web-Seiten erfasst und keine aufwendige Erfassungsprozedur für eine komplette Web Site startet. Je nach Web-Site-Größe (Anzahl der Web-Seiten im Adressraum) kann dieser Vorgang mehrere Stunden dauern. Anhand der Web-Seiten-Anzahl kann folglich der Zeitaufwand für einen ersten manuellen Durchlauf begrenzt werden.

Der Fortschritt eines manuellen Sammelvorgangs wird nach dem Start der Sammlung im unteren Teil der Eingabemaske angezeigt.

Abbildung 55 stellt die Fortschrittsanzeige der manuellen Sammlung dar.

Max. Dateigröße in MB:	<input type="text" value="5"/>
Anzahl Web-Seiten:	<input type="text" value="10"/>
System sammelt ( 4/10 ) ...	

Abb. 55: Fortschrittsanzeige manueller Sammlungen

Zur Überprüfung der gesammelten Web-Seiten können einzelne Abbilder unmittelbar nach der Sammlung abgerufen werden.

Abbildung 56 stellt das Abbild der Startseite des Fachbereichs 02 der JLU-Gießen nach einer manuellen Sammlung in der Reproduktionsumgebung des WSBS dar.

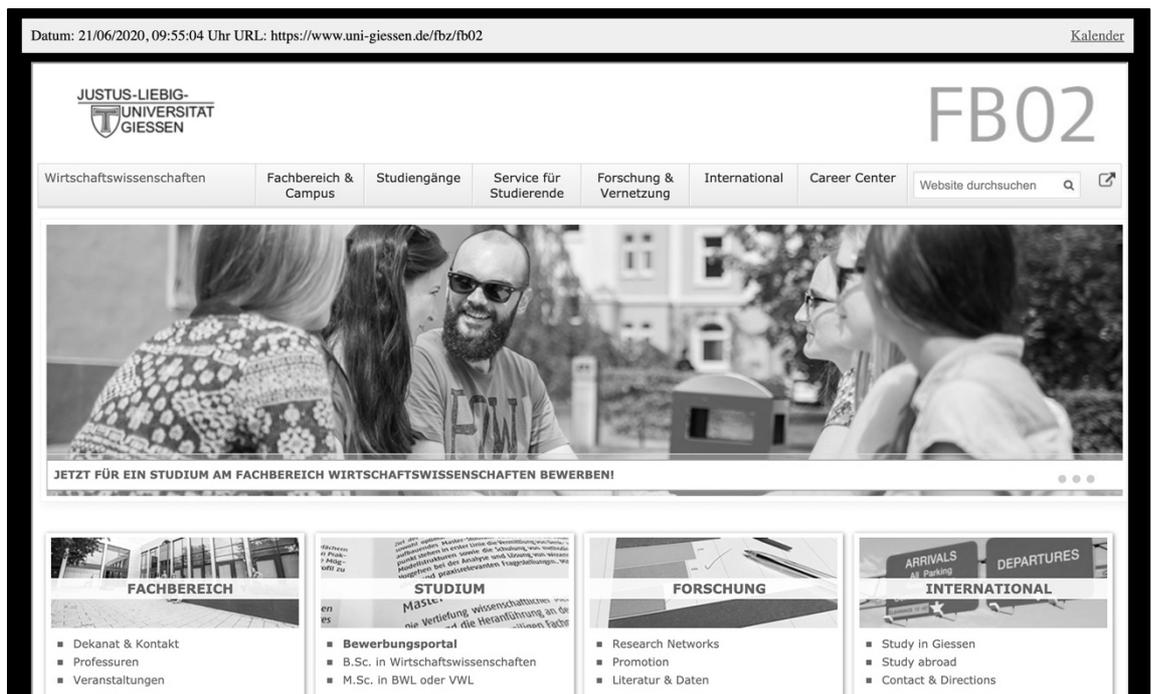


Abb. 56: Abbild der Startseite des FB02 der JLU-Gießen am 21.06.2020

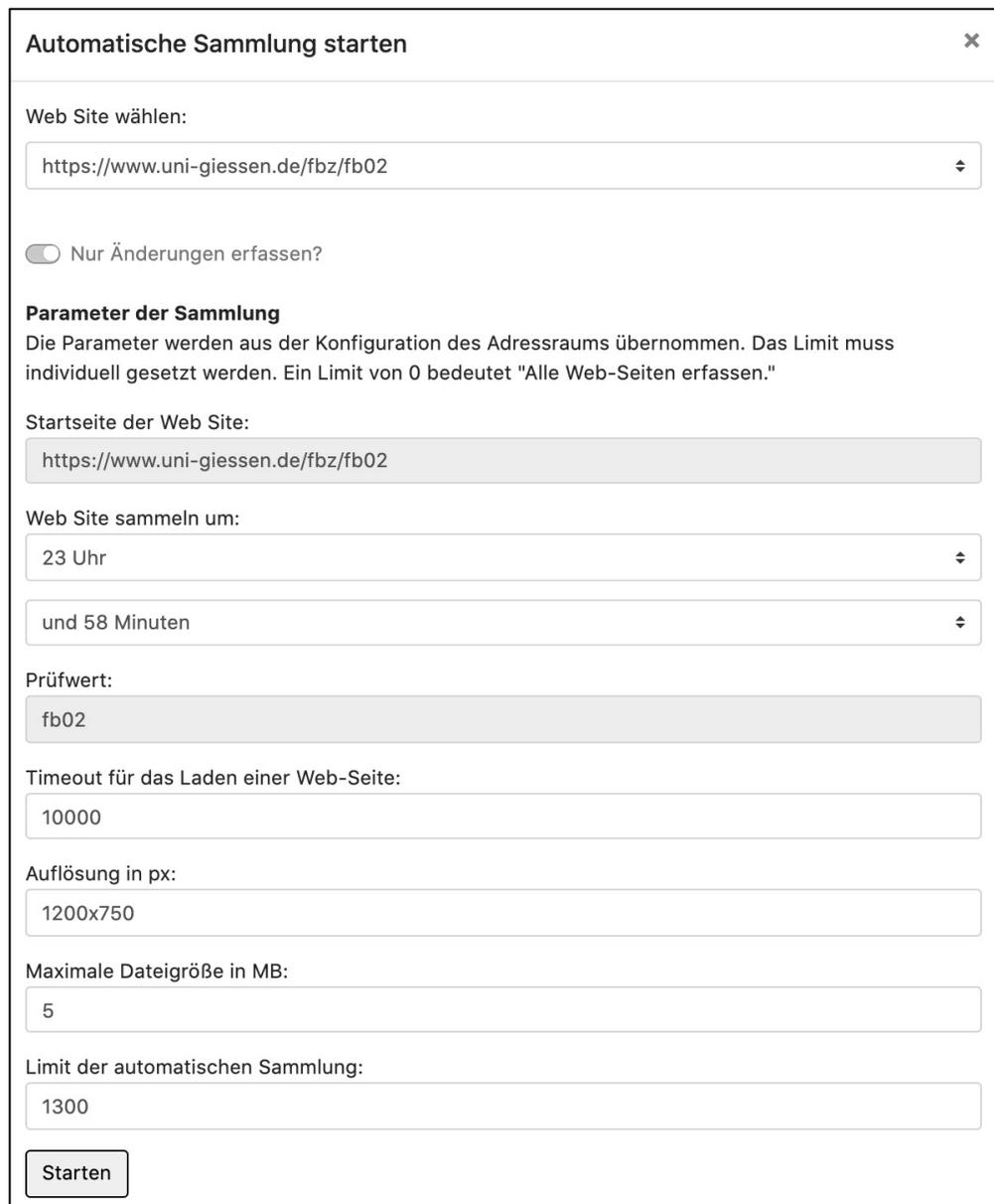
Die Betreiber des WSBS können anhand der erfassten Abbilder testen, ob Inhalte und Gestaltungselemente fehlerfrei reproduziert werden.

Sind einzelne Inhalte oder Gestaltungselemente auf einem Abbild fehlerhaft, können die Timeout-Angabe oder maximale Dateigröße nachjustiert und der Sammelprozess erneut

gestartet werden. Lassen sich keine Fehler in Inhalt und Gestaltung feststellen, kann die automatisierte Sammlung der WSB aktiviert werden.

Das Aktivieren der automatisierten Sammlung erfolgt auf einer weiteren Eingabemaske und wird im Folgenden erläutert.

Abbildung 57 stellt die Eingabemaske zum Starten der automatisierten Sammlung des WSBS dar.



The screenshot shows a web form titled "Automatische Sammlung starten" with a close button (X) in the top right corner. The form contains the following fields and controls:

- Web Site wählen:** A dropdown menu with the value "https://www.uni-giessen.de/fbz/fb02".
- Nur Änderungen erfassen?:** A toggle switch that is currently turned off.
- Parameter der Sammlung:** A section header followed by explanatory text: "Die Parameter werden aus der Konfiguration des Adressraums übernommen. Das Limit muss individuell gesetzt werden. Ein Limit von 0 bedeutet 'Alle Web-Seiten erfassen.'"
- Startseite der Web Site:** A text input field containing "https://www.uni-giessen.de/fbz/fb02".
- Web Site sammeln um:** Two stacked dropdown menus. The first contains "23 Uhr" and the second contains "und 58 Minuten".
- Prüfwert:** A text input field containing "fb02".
- Timeout für das Laden einer Web-Seite:** A text input field containing "10000".
- Auflösung in px:** A text input field containing "1200x750".
- Maximale Dateigröße in MB:** A text input field containing "5".
- Limit der automatischen Sammlung:** A text input field containing "1300".
- Starten:** A button at the bottom left of the form.

Abb. 57: Eingabemaske „Automatische Sammlung starten“

Die Eingabemaske zum Starten des automatisierten Sammelprozesses wird auf Basis der Adressraumdaten vorausgefüllt. Nachdem eine Web Site ausgewählt wurde („Web Site

wählen“), lädt das WSBS alle notwendigen Informationen zum Durchführen des Sammelvorgangs in die Eingabemaske.

Bevor die Sammlung durch Klick auf den Button „Starten“ angestoßen wird, können der Analysezeitpunkt („Web Site sammeln um“), ein Timeout, die Auflösung der Web Site und das Limit der Sammlung sowie die maximale Dateigröße angepasst werden.

Im vorliegenden Beispiel wird das Limit auf 1.300 Web-Seiten festgelegt. Wird für das Limit „0“ eingegeben, erfasst das WSBS alle Web-Seiten im Adressraum.

Nach der Aktivierung der automatischen Sammlung wird dauerhaft eine Meldung angezeigt, die signalisiert, dass der Crawler des WSBS aktiv ist.

Abbildung 58 visualisiert die Bedienoberfläche mit der entsprechenden Statusmeldung.

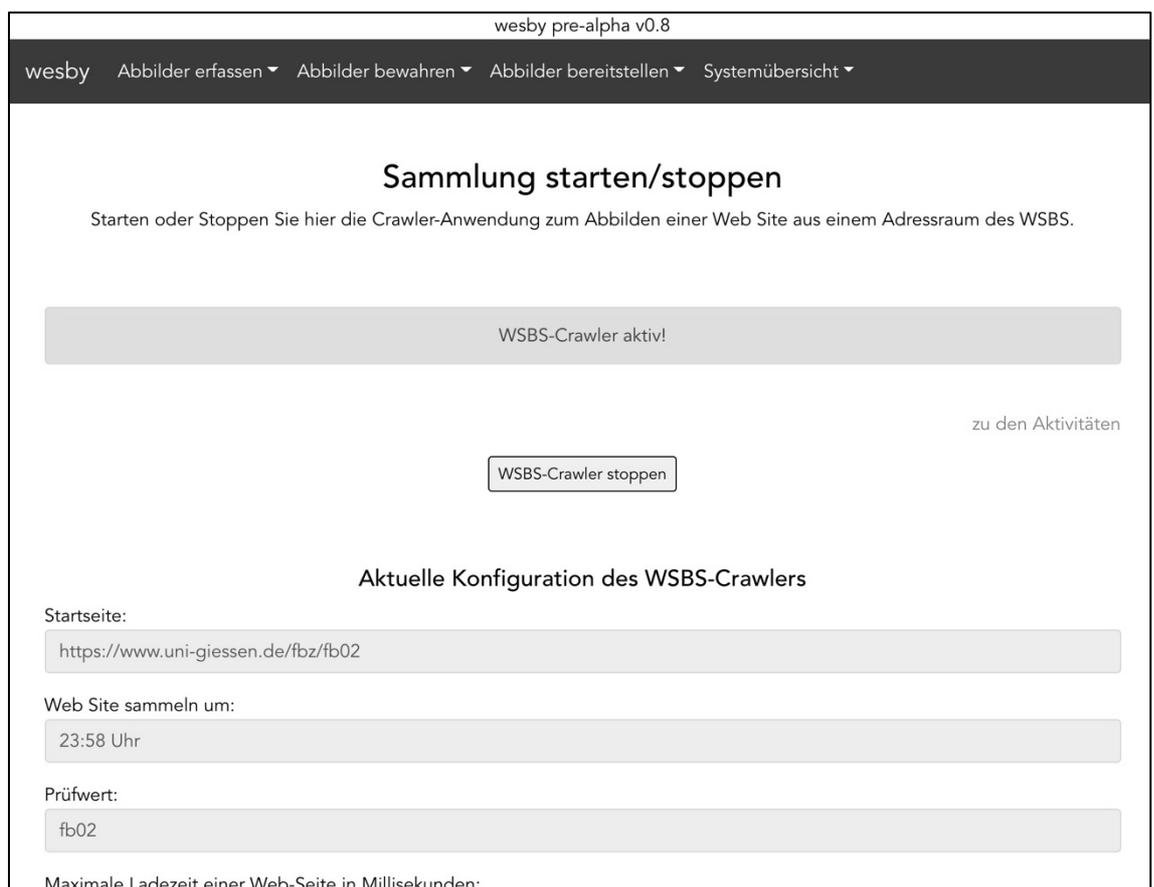


Abb. 58: Sammlung starten und stoppen

Nachdem im vorliegenden Kapitel Bedienoberflächen zum Erfassen einer WSB dargestellt wurden, widmet sich das nachfolgende Kapitel Bedienoberflächen, die zur Bewahrung von Abbildern im WSBS verwendet werden.

### 6.3.3 Bedienoberflächen zum Bewahren einer WSB

Die Bedienoberflächen zum Bewahren einer WSB werden zum Dokumentieren von Kontext-Informationen verwendet. Auf den Bedienoberflächen können nicht nur Informationen über die Hardware und Software der Web Site hinterlegt, sondern auch Informationen über die organisatorische Infrastruktur erfasst werden. Im Folgenden soll zunächst die Bedienoberfläche zur Dokumentation der technischen IT-Infrastruktur (Hard- und Software) aufgezeigt werden.

Die technische IT-Infrastruktur einer Web Site besteht aus IT-Infrastruktur-Komponenten. Eine IT-Infrastruktur-Komponente kann z. B. ein Web Server, eine Datenbank oder das Web-Content-Management-System zum Erstellen und Pflegen der Web Site sein.

Im Prototyp des WSBS stehen vordefinierte Kategorien zur Verfügung, um die einzelnen IT-Infrastruktur-Komponenten einer Web Site zu klassifizieren:<sup>420</sup>

- Web-Content-Management-System (WCMS)
- Web Server
- Datenbanksystem
- Betriebssystem
- Virtualisierungs-Software
- Computing-Hardware
- Storage-Hardware

Auf Basis der Kategorien können technische Infrastruktur-Komponenten angelegt und verwaltet werden.

Eine Übersichtseite im WSBS listet alle technischen Infrastruktur-Komponenten auf.

---

<sup>420</sup> Vgl. z. B. Rudolph, Simone: Servicebasierte Planung und Steuerung der IT-Infrastruktur im Mittelstand: ein Modellansatz zur Struktur der IT-Leistungserbringung, a .a .O., S. 16 f.

Abbildung 59 stellt die Übersicht zur Verwaltung der technischen IT-Infrastruktur dar.

wesby pre-alpha v0.8

wesby Abbilder erfassen ▾ Abbilder bewahren ▾ Abbilder bereitstellen ▾ Systemübersicht ▾

## Technische IT-Infrastruktur

Dokumentieren Sie IT-Infrastruktur-Komponenten, die zum Betrieb einer Web Site verwendet werden oder wurden.

[IT-Infrastruktur-Komponente hinzufügen](#)

Löschen deaktiviert

Alle IT-Infrastruktur-Komponenten oder IT-Infrastruktur-Komponenten pro Adressraum bzw. Web Site anzeigen:

Web Site "https://www.uni-giessen.de/fbz/fb02" ▾

Komponente	Kategorie	Einsatzzeitraum	Aktiv?
<b>Plone 4</b> Beschreibung: Dezentrales Web-Content-Management-System zum Betreiben der Web Site des Fachbereichs 02 an der JLU-Gießen. Passende Abbilder im Einsatzzeitraum öffnen: <a href="https://www.uni-giessen.de/fbz/fb02">https://www.uni-giessen.de/fbz/fb02</a>	WCMS	09.04.2014 – unbekannt	Ja <a href="#">Bearbeiten</a>

Abb. 59: Technische IT-Infrastruktur im WSBS verwalten

Die technischen IT-Infrastruktur-Komponenten im WSBS lassen sich nachträglich bearbeiten und bei Bedarf löschen. Nur deaktivierte IT-Infrastruktur-Komponenten können gelöscht werden.

Ergänzend zur Dokumentation von technischen IT-Infrastruktur-Komponenten können Mitarbeiter als organisatorische Infrastruktur erfasst werden.

Zur Dokumentation der Mitarbeiter muss, neben einem Tätigkeitszeitraum, eine oder mehrere Rollen ausgewählt werden. Zur Auswahl stehen folgende Rollen:

- Web-Site-Verantwortlicher
- Web-Site-Administrator
- Web-Site-Designer
- Web-Site-Redakteur

Web-Site-Verantwortliche sind innerhalb der Organisation für den Betrieb der Web Site verantwortlich, führen aber nicht zwingend operative Aufgaben aus. Web-Site-Administratoren sind dagegen mit der operativen Bereitstellung der Betriebsumgebung einer Web Site betraut. Web-Site-Designer konzipieren und realisieren die Gestalt einer organisationalen Web Site und Web-Site-Redakteure sind ausschließlich mit dem Erstellen und Publizieren von Inhalten beschäftigt.

Vergleichbar mit der Bedienoberfläche zur Administration der technischen IT-Infrastruktur ist auch die organisatorische Infrastruktur verwaltbar.

Abbildung 60 stellt die Bedienoberfläche zum Dokumentieren der Mitarbeiter dar.

wesby pre-alpha v0.8

wesby Abbilder erfassen ▾ Abbilder bewahren ▾ Abbilder bereitstellen ▾ Systemübersicht ▾

## Mitarbeiter als organisatorische Infrastruktur dokumentieren

Dokumentieren Sie hier das Personal, das zum Betrieb einer Web Site eingesetzt wird oder wurde.

Mitarbeiter hinzufügen

Löschen deaktiviert

Alle Mitarbeiter oder Mitarbeiter pro Adressraum bzw. Web Site anzeigen:

Web Site "https://www.uni-giessen.de/fbz/fb02" ▾

Name	Rolle	Tätigkeitszeitraum	Aktiv?
<b>Maximilian Mustermann</b>	Web-Site-Designer Web-Site-Redakteur	01.06.2015 bis 31.05.2020	Ja

Bearbeiten

Beschreibung:  
Mitarbeiter zum Erstellen und Pflegen von Web-Seiten am Dekanat des Fachbereichs 02.

Mitarbeiter der Organisation zum Betrieb der Web Site:  
**https://www.uni-giessen.de/fbz/fb02**

Abb. 60: Dokumentation von Mitarbeitern im WSBS

Die Dokumentation der Mitarbeiter lässt sich im WSBS genau wie IT-Infrastruktur-Komponenten bearbeiten und bei Bedarf löschen.

Neben dem Dokumentieren von technischen und organisatorischen Infrastruktur-Komponenten umfasst die Bewahrung im WSBS auch das Verwalten der kopierten Abbilder.

Abbilder können nach ihrer Erfassung im WSBS gesucht und für die Bereitstellung in einer WSB gesperrt, bearbeitet oder gelöscht werden.

Abbildung 61 zeigt die Bedienoberfläche zum Suchen und Verwalten einzelner HTML-Dateien.

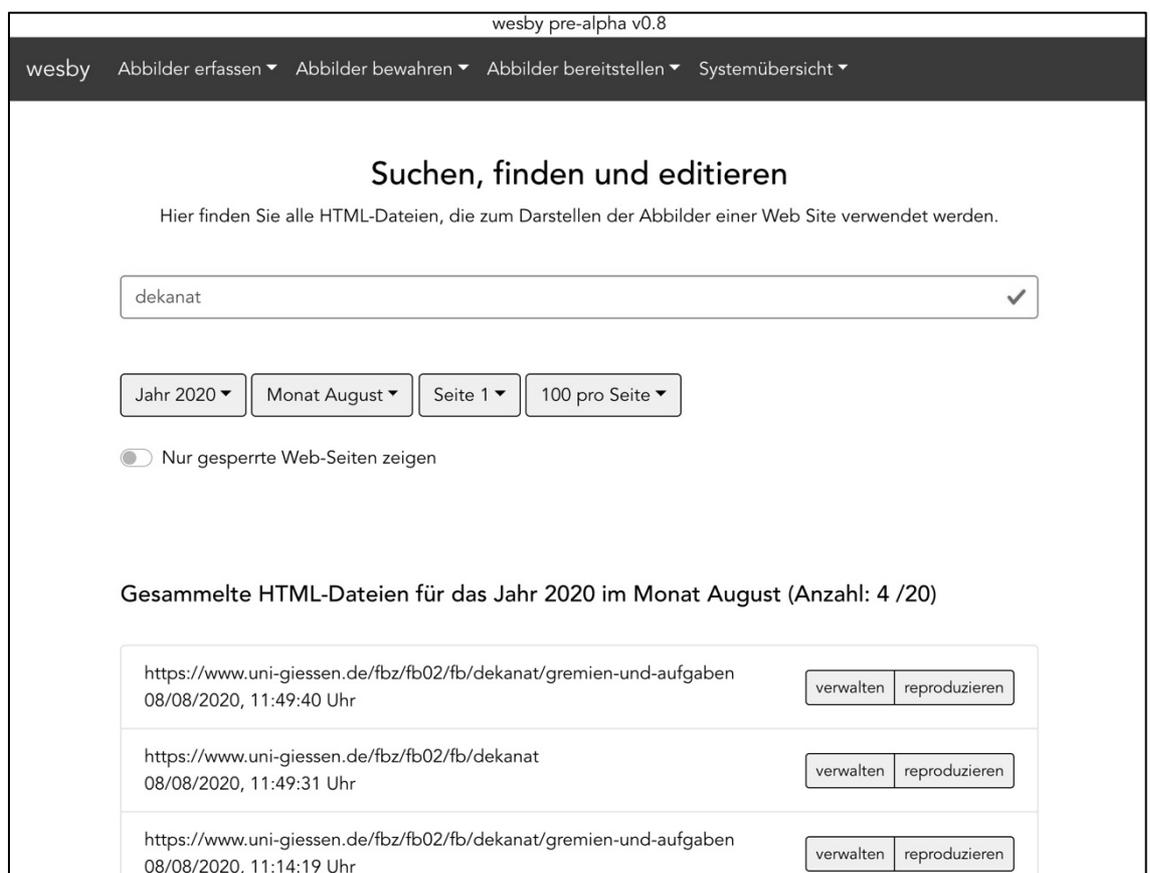


Abb. 61: Abbilder suchen und verwalten

Die erfassten HTML-Dateien werden nach ihren Erfassungszeitpunkten aufgelistet. Damit rechtskritische oder sensible Informationen nicht unveränderbar im WSBS konserviert bleiben, besteht die Möglichkeit, einzelne Abbilder zu sperren oder Inhalte der Abbilder aus dem Datenbestand zu löschen. Über den „verwalten“-Button können einzelne Abbilder geöffnet und nachträglich editiert werden.

Abbildung 62 zeigt die Bedienoberfläche zum Verwalten der Abbilder im WSBS.

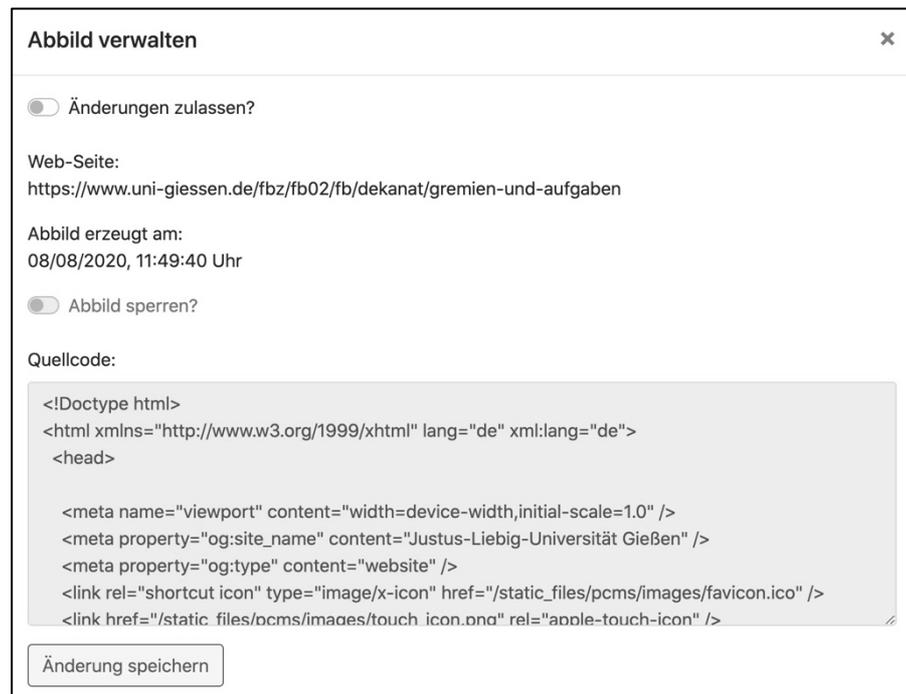


Abb. 62: Verwalten von Abbildern im WSBS

Einzelne Abbilder im WSBS lassen sich sperren und entsperren. Bei Bedarf kann auch der Quellcode einer Web-Seite bearbeitet und im Abbild gespeichert werden.

Das nachträgliche Editieren des Quellcodes sollte zur Wahrung der Echtheit eines einzelnen Abbildes vermieden werden. Bestehen allerdings rechtliche Vorgaben, die ein nachträgliches Editieren notwendig machen, können Veränderungen mit dem WSBS umgesetzt werden.

Nachdem im vorliegenden Kapitel ausgewählte Bedienoberflächen zum Bewahren einer WSB dargestellt wurden, widmet sich das nachfolgende Kapitel den Bedienoberflächen, die zur Bereitstellung von Abbildern verwendet werden.

### 6.3.4 Bedienoberflächen zum Bereitstellen einer WSB

Die Bedienoberflächen zum Bereitstellen einer WSB sind in eine Bedienoberfläche zur Konfiguration der Bereitstellungsumgebung und in Bedienoberflächen zum Erleben der WSB gegliedert. Im Folgenden soll zunächst die Bedienoberfläche zum Konfigurieren der Bereitstellungsumgebung erläutert werden.

Abbildung 63 visualisiert die Bedienoberfläche zur Konfiguration der Bereitstellungsumgebung im WSBS.

wesby pre-alpha v0.8

wesby Abbilder erfassen ▾ Abbilder bewahren ▾ Abbilder bereitstellen ▾ Systemübersicht ▾

## Web-Oberfläche des Systems konfigurieren

Konfigurieren Sie hier die Web-Oberfläche zum Erleben der WSB

WSB veröffentlichen?

Ja ▾

Kalender-Ansicht aktivieren?

Ja ▾

Slide-Show aktivieren?

Nein ▾

Kontext-Informationen-Anzeige aktivieren?

Nein ▾

Suchfunktion aktivieren?

Nein ▾

Konfiguration speichern

Abb. 63: Konfiguration der Web-Oberfläche zum Erleben der WSB

Betreiber des WSBS können die Veröffentlichung der WSB aktivieren oder deaktivieren. Wird die WSB für die Öffentlichkeit bereitgestellt, kann die Lebensgeschichte einer Web Site im WWW abgerufen werden. Ist die Veröffentlichung der WSB deaktiviert, kann nur nach einem erfolgreichen Login auf die Reproduktionsumgebung zugegriffen werden.

Neben der Slide-Show-basierten Darstellung der Evolution einer Web Site, kann eine Kalender-Ansicht, die Publikation von Kontext-Informationen und eine Suchfunktion aktiviert werden. Diese Funktionalitäten sind jedoch noch kein Bestandteil des Prototyps. Der Prototyp fokussiert sich zunächst auf die *originalgetreue* Reproduktion.

Die Bedienoberfläche zum Erleben einer WSB reproduziert die Web-Seiten im Web Browser des Anwenders und macht Abbilder originalgetreu nutzbar.

Abbildung 64 zeigt einen Ausschnitt der Bedienoberfläche zum Erleben eines Abbildes.

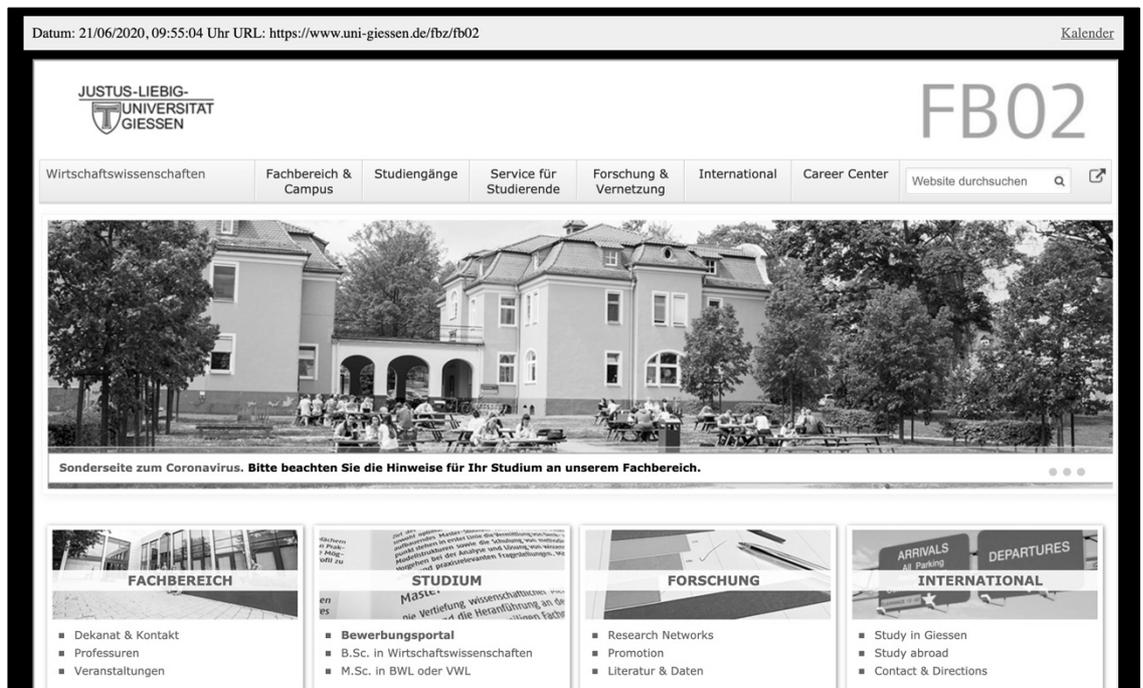


Abb. 64: Abbild einer Web-Seite des FB02 mit 1200 x 750 Pixel (1/2)

In der Kopfzeile der Bedienoberfläche wird der Zeitstempel der Erstellung ebenso wie die Web-Adresse der abgebildeten Web-Seite angezeigt.

Abbildung 65 zeigt die Kopfzeile in vergrößerter Form.



Abb. 65: Kopfzeile Reproduktionsumgebung

Auf der rechten Seite der Kopfzeile ist ein Menüpunkt zum Darstellen des Kalenders platziert. Unterhalb der Kopfzeile wird das Abbild in einem Darstellungsbereich wiederhergestellt und erlebbar gemacht.

Im Darstellungsbereich sind sämtliche Elemente eines Abbildes nutzbar. Wird z. B. im Navigationsmenü des in Abbildung 64 dargestellten Abbildes der Menüpunkt „Fachbereich & Campus“ angewählt, öffnet sich ein passendes Abbild der Web-Seite.

Mit Klick auf den Menüpunkt „Fachbereich & Campus“ wird ein Abbild der „Fachbereich & Campus“-Web-Seite geladen. Der Erfassungszeitpunkt der „Fachbereich & Campus“-Web-Seite orientiert sich am Erfassungszeitpunkt des zuvor geöffneten Abbildes.

Abbildung 66 demonstriert das Abbild der „Fachbereich & Campus“-Web-Seite.

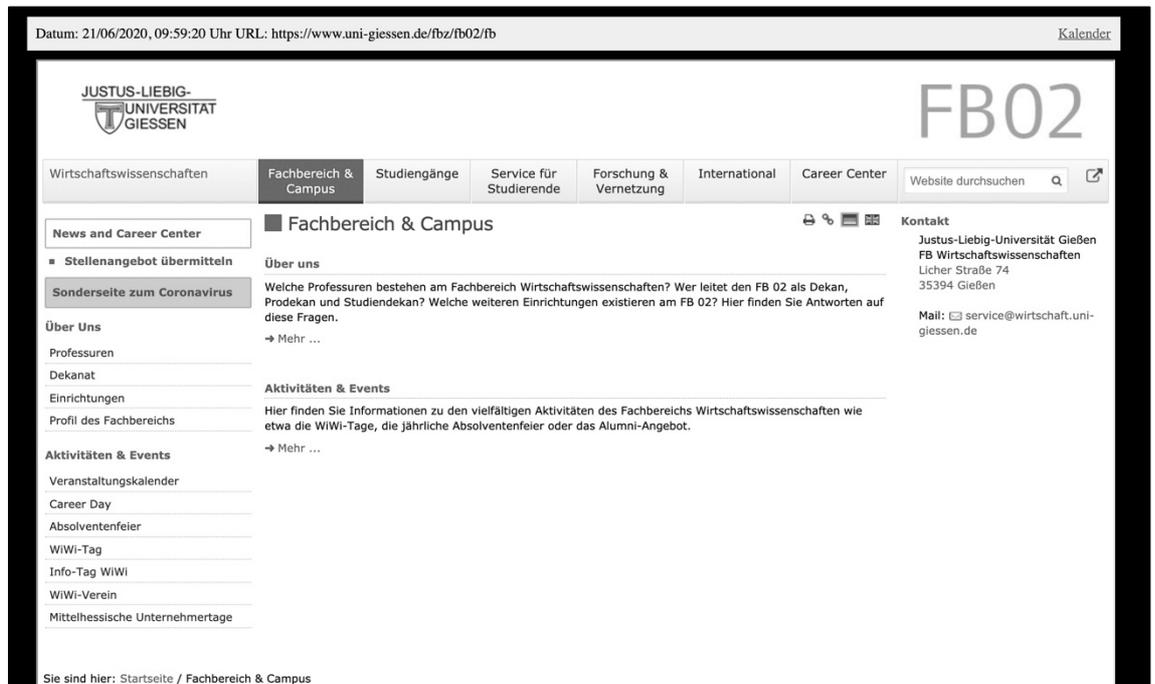


Abb. 66: Abbild einer Web-Seite des FB02 mit 1200 x 750 Pixel (2/2)

Die in Abbildung 66 visualisierte „Fachbereich & Campus“-Web-Seite wurde am 21.06.2020, um 09:59 Uhr und 20 Sekunden abgebildet. Die zuvor geöffnete Web-Seite – die Startseite des Fachbereichs 02 – am 21.06.2020, um 09:55 Uhr und 4 Sekunden.

Neben dem Nutzen von Links im Quellcode eines Abbildes, kann das Abrufen von Abbildern einer spezifischen Web-Seite auch über eine Kalenderansicht erfolgen. Die Kalenderansicht wird durch einen Klick auf den „Kalender“-Button im rechten Bereich der Kopfzeile geöffnet. Der Kalender stellt Erfassungszeitpunkte einer Web-Seite strukturiert nach Jahren und Monaten dar.

Auf der folgenden Seite wird die Kalenderansicht für die Web Site des Fachbereichs 02 der JLU-Gießen dargestellt.

Ein Ausschnitt der Kalenderansicht wird in Abbildung 67 visualisiert.

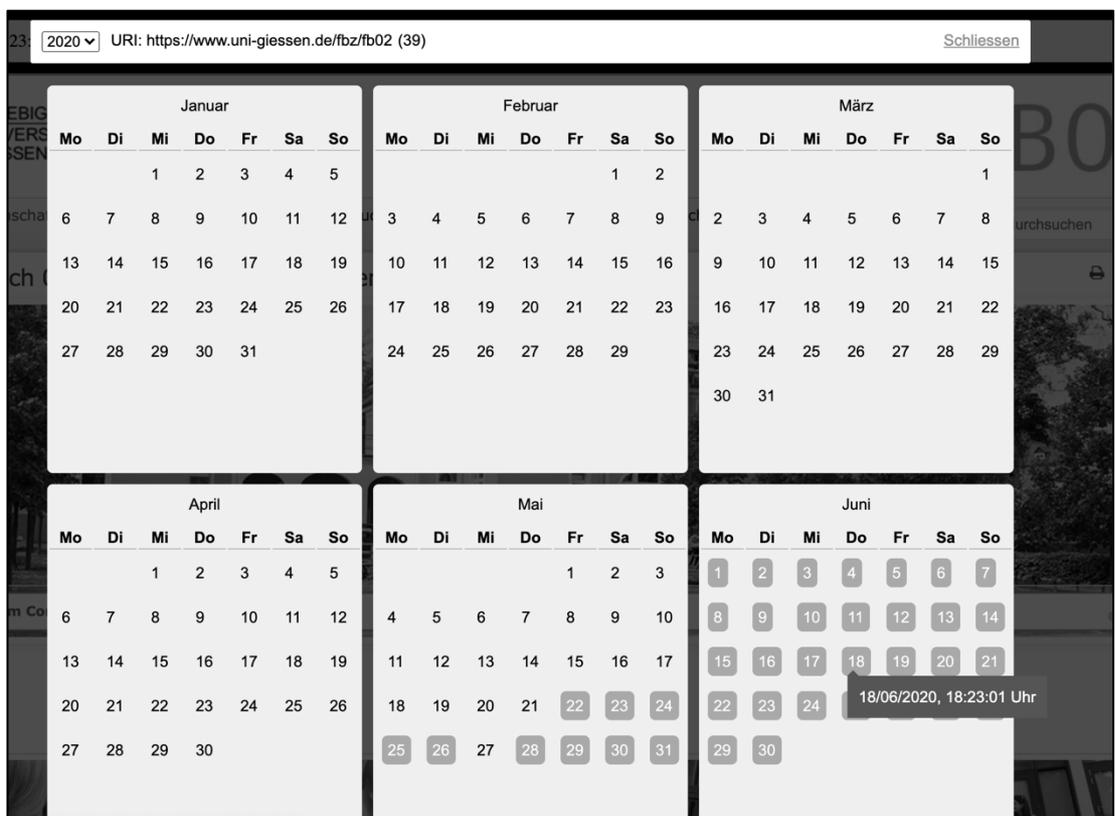


Abb. 67: Ausschnitt der Kalenderansicht

Das WSBS öffnet den Kalender für das Jahr, in dem das zugrunde liegende Abbild erfasst wurde. In der Kopfzeile kann anschließend ein Jahreswechsel vorgenommen und Abbilder zu unterschiedlichen Zeitpunkten abgerufen werden.

Abbildung 68 stellt die Kopfzeile in vergrößerter Form dar.



Abb. 68: Kopfzeile Kalenderansicht



Der Vergleich der Häufigkeit von Sammlungen bzw. Abbildern im Internet-Archiv mit den Sammlungen bzw. Abbildern im Prototyp des WSBS veranschaulicht sehr deutlich, dass nur durch eine dedizierte Systemlösung, ein WSBS, die vollständige Lebensgeschichte einer Web Site (WSB) umfänglich abgebildet werden kann. Global agierende Sammellösungen können diesen Umfang nicht leisten.

Nachdem im vorliegenden Kapitel die Bedienoberfläche zum Bereitstellen der Lebensgeschichte einer Web Site beschrieben und ein Vergleich mit dem Internet-Archiv hergestellt wurde, zeigt Kapitel 6.3.5 abschließend weitere Anwendungen des Prototyps auf.

### 6.3.5 Weiteres Anwendungsbeispiel

Da Web Sites unterschiedlichste technische Ausprägungen, Inhalte und Gestaltungselemente aufweisen, soll das vorliegende Kapitel den Einsatz des Prototyps für eine weitere Web Site zeigen. Vorab soll an dieser Stelle erwähnt werden, dass die in diesem Kapitel gezeigten Abbilder zur wissenschaftlichen Auswertung in einer isolierten technischen Umgebung erzeugt wurden. Einzelne Web-Seiten wurden für einen Nachweis der Funktionalität des WSBS reproduziert. Das WSBS hat dabei die Funktion einer Reproduktionsumgebung bereitgestellt und keine dauerhafte Speicherung, Veröffentlichung oder Analyse umgesetzt, um die Eigentumsrechte, insbesondere das Urheberrecht des Web-Site-Betreibers zu wahren.<sup>422</sup>

Als Anwendungsbeispiel wird im Folgenden die Web Site der BMW AG herangezogen. Die Web Site der BMW AG entsteht aus einer komplexen Kombination von Inhalts- und Gestaltungselementen, die das WSBS reproduzieren muss, um eine einzelne Web-Seite originalgetreu anzuzeigen.

Abbildung 70 stellt ein Abbild der Startseite der BMW-Web-Site beispielhaft dar.

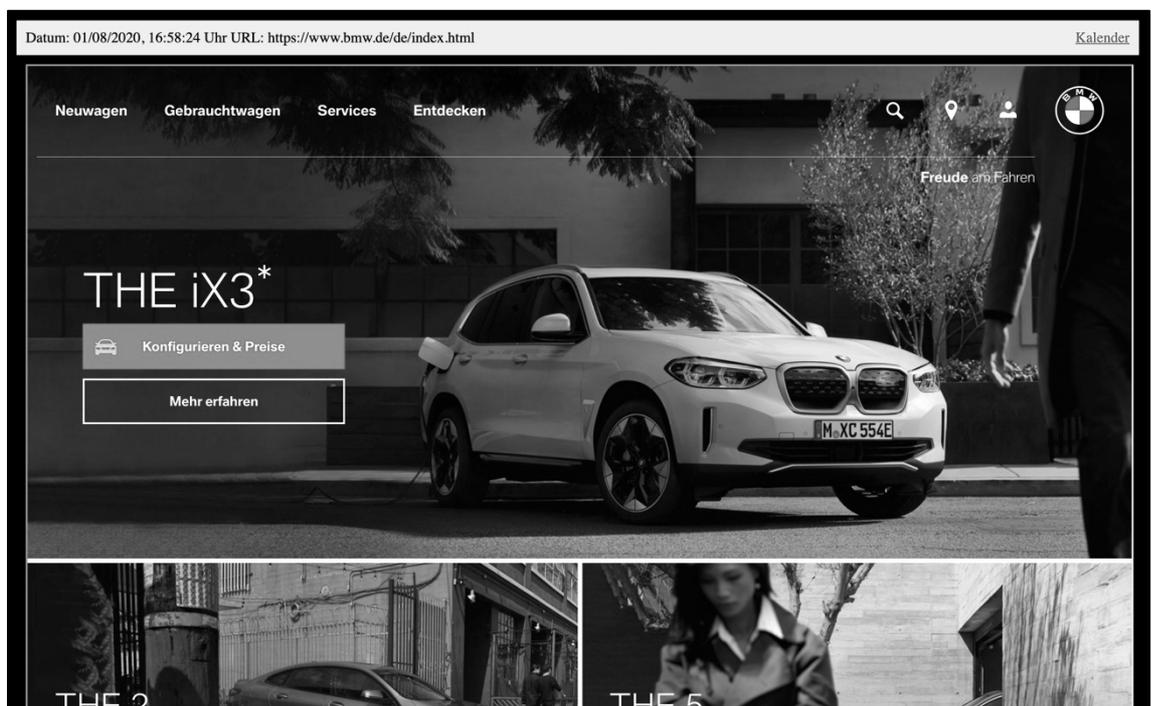


Abb. 70: Demo-Abbild von <https://www.bmw.de>, 01.08.2020

<sup>422</sup> Vgl. Brown, Adrian: Archiving websites: a practical guide for information management professionals, a. a. O., S. 147 ff. und Pastukhov, Oleksandr; Dobrova, Milena (Hrsg.): Legal issues surrounding digital archiving, in: Digital Archives: Management, use and access, London: Facet Publishing 2018, S. 74 ff.

Wie mit Abbildung 70 visualisiert wird, kann das WSBS die Startseite der BMW-Web-Site ohne grafische Fehler reproduzieren. Bilder und Funktionen werden geladen und Verlinkungen zu weiteren Abbildern integriert. Mit Klick auf den Menüpunkt „Neufahrzeuge“ öffnet sich die zugehörige Web-Seite in der Reproduktionsumgebung des WSBS.

Das Abbild der Neufahrzeuge-Web-Seite ist durch umfangreiche Bilder-Sammlungen charakterisiert. Einzelne Bilder der Sammlung werden auf der original Web-Seite erst nach Nutzerinteraktionen geladen und angezeigt. Ein Besucher muss sich zunächst im Web Browser nach unten bewegen, bevor weitere Bilder geladen werden. Diese Funktion kann vom WSBS nachgestellt und in der Reproduktionsumgebung wiedergegeben werden. Durch den hybriden Ansatz wird die Nutzerinteraktion zunächst im HWB des WSBS simuliert und anschließend von der Recorder-Software in die Reproduktionsumgebung übertragen. Das Replay-Skript sorgt anschließend für das Nachladen der Inhalte, wenn sich ein Nutzer in der Reproduktionsumgebung zum Ende der Web-Seite bzw. des Abbildes bewegt.

Abbildung 71 stellt die Bilder-Sammlung der Neufahrzeuge auf der BMW-Web-Site beispielhaft dar.

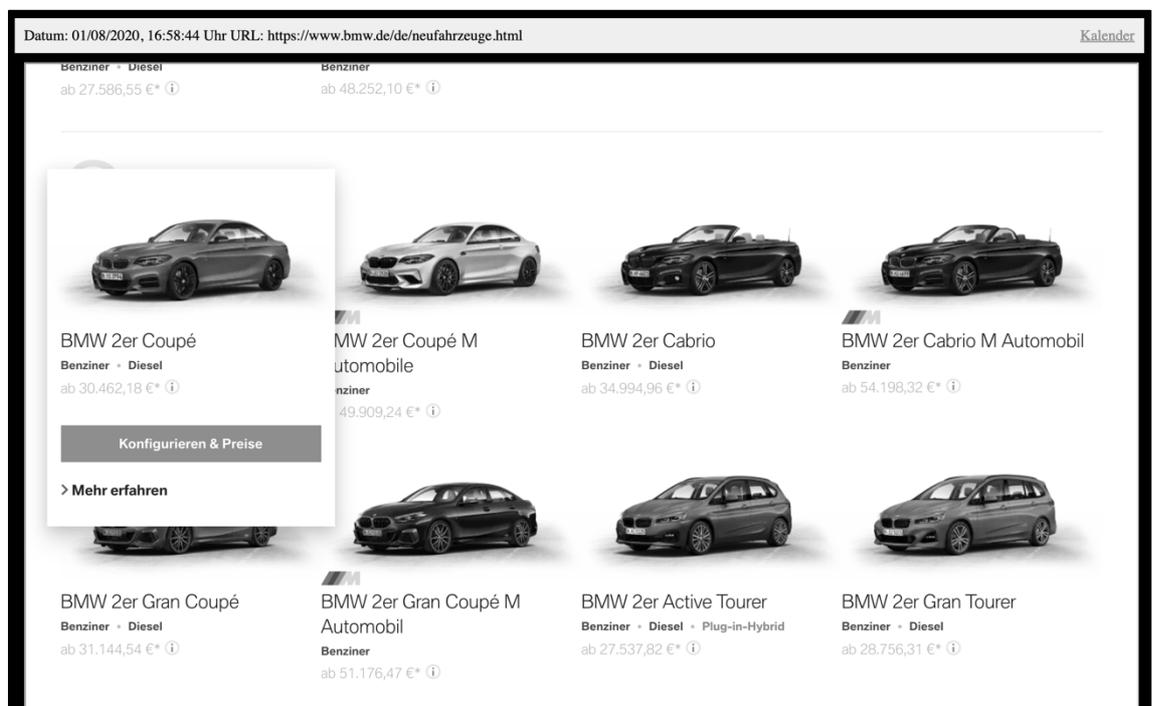


Abb. 71: Demo-Abbild von <https://www.bmw.de/neufahrzeuge>, 01.08.2020

Ergänzend zu der vollständigen Wiedergabe von komplexen Bild-Beständen, bietet das WSBS die Möglichkeit, mobile Versionen einer Web Site zu reproduzieren.

Um eine mobile Version einer Web Site zu erfassen, muss bei der Konfiguration des Erfassungsprozesses eine Auflösung für mobile Endgeräte hinterlegt werden (vgl. Abbildung 57). Eine typische Auflösung sind z. B. 375 x 812 Pixel.

Abbildung 72 demonstriert beispielhaft die mobile Version der Startseite der BMW-Web-Site mit einer Auflösung von 375 x 812 Pixel.

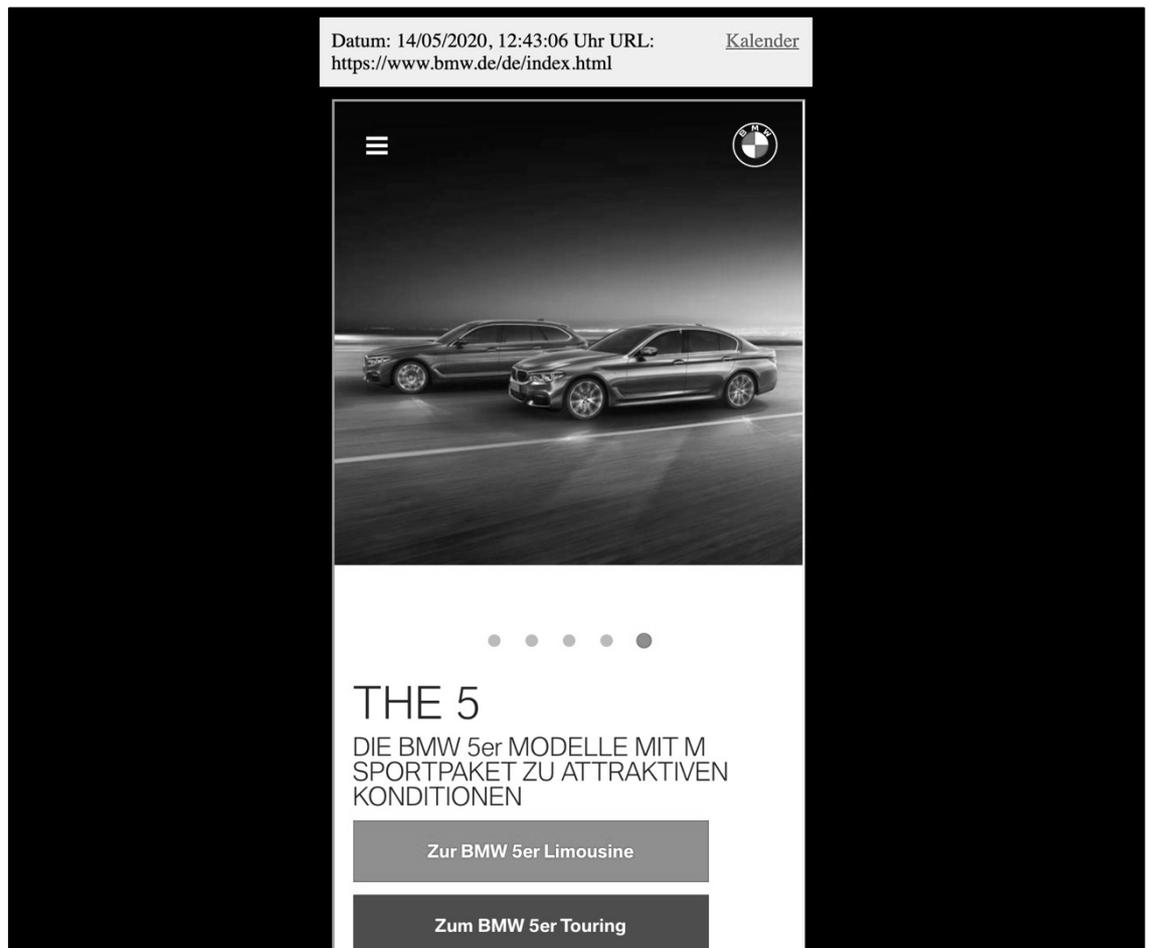


Abb. 72: Demo-Abbild einer Web-Seite mit 375 x 812 Pixel

In Abbildung 72 wird die Reproduktion der mobilen Web Site der BMW AG am 14.05.2020 beispielhaft visualisiert. Die Web-Seite wurde mit einer Auflösung von 375 x 812 Pixeln im WSBS erfasst und wiederhergestellt.

Bei der Wiederherstellung eines Abbildes passt sich die Reproduktionsumgebung automatisch an die Auflösung des Abbildes an und ermöglicht so das Erleben originalgetreuer, mobiler Versionen einer Web Site.

## 7 Zusammenfassung, Bewertung und Ausblick

Kulturgüter werden längst in Form von digitalen Bildern, Texten, Fotos, Videos, E-Mails, Spielen oder Web Sites erzeugt.<sup>423</sup> Als Vorreiter internationaler Institutionen, die sich vor allem mit der Bewahrung des Kulturgutes „Web Site“ beschäftigen, kann das Internet-Archiv angeführt werden. Das Internet-Archiv wurde als digitales Gedächtnis zur Sammlung und Aufbewahrung von Publikationen aus dem World Wide Web gegründet und sammelt bereits mehr als eine Milliarde Web-Seiten pro Woche.<sup>424</sup>

Im Gegensatz zur Sammel-Motivation des Internet-Archivs, ist bei Unternehmen aus Wirtschaft und Verwaltung kaum bzw. keinerlei Anstrengung zu erkennen, die vergangenen Zustände der *eigenen* Organisations-Web-Site zu konservieren.<sup>425</sup> Erhaltungsmaßnahmen seitens der Unternehmen sind jedoch zwingend notwendig, wenn die Zustände der Web Site vollständig und verlässlich für die Nachwelt dokumentiert werden sollen.

Global agierende Initiativen wie das Internet-Archiv haben bis heute keine Möglichkeit, sämtliche Zustände aller im WWW veröffentlichten Web Sites zu sichern und diese für historische, nostalgische oder betriebswirtschaftliche Auswertungen aufzubewahren. Eine Vielzahl der Web Sites ist in den Sammlungen der Web-Archive fehlerhaft, unvollständig oder überhaupt nicht vertreten.<sup>426</sup> Bei der Übertragung von Web-Seiten in die Infrastruktur eines Web-Archivs gehen nicht selten Bilder, Texte, Funktionen oder andere Inhalte verloren.<sup>427</sup>

Eine System-Lösung, die auf das oben geschilderte Problem reagiert und sämtliche Zustände einer Web Site langfristig und verlässlich sichert, wurde in der vorliegenden Arbeit hergeleitet und prototypisch realisiert, das *Web-Site-Biographie-System (WSBS)*. Das WSBS hat die Aufgabe, die „Lebensgeschichte“ einer Web Site – die Web-Site-Biographie – möglichst vollständig und fehlerfrei zu konservieren. Die Zustände der Web Site werden durch das System originalgetreu erlebbar gemacht.

---

423 Vgl. Gesellschaft für Informatik e. V. (Hrsg.): Digitales Kulturerbe, a. a. O., 28.03.2017.

424 Vgl. Lobe, Adrian: Wir sammeln eine Milliarde Seiten pro Woche, a. a. O., S. 13.

425 Ein weltweiter Überblick über Archivierungsinitiativen findet sich in Schwenck, Rebecca: Langzeitarchivierung digitaler Informationen: Ein internationaler Überblick über Organisation und Projektarbeit, a. a. O., S. 26 ff.

426 Vgl. z. B. Masanès, Julien: Web archiving, a. a. O., S. 20 oder Brown, Adrian: Archiving websites: a practical guide for information management professionals, a. a. O., S. 28 oder Brügger, Niels: Web historiography and Internet Studies: Challenges and perspectives, a. a. O., S. 5 f. oder Haber, Peter: Digital past: Geschichtswissenschaft im digitalen Zeitalter, a. a. O., S. 61.

427 Vgl. Masanès, Julien: Web archiving, a. a. O., S. 15 oder Hessischer Rundfunk (Hrsg.): Gegen den Gedächtnisverlust: Das Archiv im digitalen Zeitalter, a. a. O., 05.05.2017.

Zur systematischen Entwicklung eines Prototyps wurden zunächst die Eigenschaften einer organisationalen Web Site definiert und beschrieben. Die Beschreibungen der Eigenschaften fokussieren sich vor allem auf Herausforderungen, die im Kontext der digitalen Erhaltung entstehen und von einem WSBS gelöst werden müssen. Aufbauend auf der Definition und Beschreibung der Eigenschaften folgt eine Anforderungsanalyse. Die Anforderungsanalyse umfasst fachliche, organisatorische und technische Anforderungen, die zur Realisierung des Prototyps umgesetzt werden. Unter Berücksichtigung der Anforderungen wird das WSBS zunächst aus einer technisch-konzeptionellen Perspektive entworfen. Durch eine Zerlegung in abgeschlossene Systemmodule entsteht ein strukturierter Bauplan. Die Summe der Systemmodule stellt die ganzheitliche Systemlösung zum Erfassen, Bewahren und Bereitstellen der WSB dar und bildet das konzeptionelle Fundament für die Realisierung des Prototyps.<sup>428</sup>

Der Prototyp des WSBS ist durch eine Kombination von Verfahren zur Erfassung der Inhalte und Gestaltungselemente einer Web Site ausgezeichnet. Durch eine Kombination des Crawler-basierten und transaktionalen Erfassungsverfahrens können Nutzeraktivitäten simuliert und sämtliche Inhalte in ein Abbild einer Web-Seite übertragen werden. Dieses *hybride* Erfassungsverfahren wurde mit Hilfe eines Headless Web Browsers, einer Recorder-Software und einem Replay-Skript umgesetzt und in einer Crawler-Anwendung konsolidiert.

Die Crawler-Anwendung fokussiert das Kopieren öffentlicher Web-Seiten. Das Kopieren von Web-Seiten, die in abgeschlossenen Bereichen einer Web Site vorgehalten werden, wird nicht durchgeführt. Sicherheitsabfragen erschweren bzw. verhindern das automatisierte Öffnen abgesicherter Bereiche einer Web Site. Das IT-System-basierte Lösen dieser Sicherheitsabfragen setzt nicht nur erhebliche Rechenleistungen voraus, sondern ist mit erheblichem Zeitaufwand verbunden. Ein effizientes und sequentielles Abfragen einer kompletten Web Site ist daher oftmals nicht möglich. Aus diesem Grund wird im vorliegenden Prototyp von der Erfassung abgesicherter Inhalte zunächst abgesehen.

Forschungsbedarfe lassen sich aus der Herleitung und Bestimmung der Anforderungen an ein WSBS ableiten. Ein lehrbuchmäßiges Vorgehen zur Definition und Evaluation von Anforderungen inkludiert Befragungen und Bewertungen der Nutzer, die ein entsprechendes Anwendungssystem verwenden sollen.<sup>429</sup> Das Einbeziehen der Nutzer setzt im Regelfall voraus, dass Kenntnisse über die Anwendungsdomäne des IT-Systems vorhan-

---

428 Vgl. Ludewig, Jochen; Lichter, Horst: Software Engineering: Grundlagen, Menschen, Prozesse, Techniken, a. a. O., S. 400.

429 Vgl. Ludewig, Jochen; Lichter, Horst: Software Engineering: Grundlagen, Menschen, Prozesse, Techniken, a. a. O., S. 366 f.

den sind. Da ein WSBS einen Anwendungsbereich innerhalb eines Unternehmens umfasst, der bis heute kaum bzw. überhaupt nicht erschlossen wurde, ist das Einbeziehen der Nutzer bei einem initialen Systementwurf herausfordernd. Kenntnisse über die Anwendungsdomäne sind im Regelfall nicht vorhanden. Eine sich dieser Arbeit anschließende Forschungsfrage kann daher auf der Analyse bzw. Bewertung der hier formalisierten Anforderungen aufbauen. Nachdem ein WSBS über einen adäquaten Betriebszeitraum implementiert und genutzt wurde, kann rückwirkend eine Bewertung der Anforderungen erfolgen.

Weitere Forschungsarbeiten könnten sich folgenden Fragen widmen: Wird die Web Site einer Organisation mit einem WSBS über Jahre dauerhaft und vollständig erfasst? Kann das WSBS als Backup-System zum Bereitstellen von Web Sites verwendet werden?

Besonders der Einsatz als Backup-System erscheint im Kontext eines „Hacker-Angriffes“ und damit verbundener Systemausfälle relevant: Wenn die IT-Infrastruktur einer Organisation aus Sicherheitsgründen abgeschaltet wird, ist auch die Organisations-Web-Site betroffen. Dokumente, die Kunden oder Mitarbeitern über die Web Site bereitgestellt werden, sind nicht mehr verfügbar. Eine unmittelbare Beeinflussung geschäftlicher Aktivitäten ist die Folge. Wird auf Basis eines WSBS eine vollständige Kopie der organisationalen Web Site vorgehalten, kann eine Kopie der Web Site auf Knopfdruck reproduziert und sämtliche Dokumente weiterhin zur Verfügung gestellt werden. Voraussetzung für die Bereitstellung einer solchen Übergangs-Web-Site ist die Auslagerung eines WSBS auf organisationsexterne und performante IT-Infrastrukturen, die unabhängig von der restlichen Organisations-IT betrieben werden. Das WSBS muss zudem eine Reproduktionsumgebung bereitstellen, deren Performance und Leistungsfähigkeit mit produktiven Web Sites vergleichbar ist. Die Weiterentwicklung eines WSBS zu einer jederzeit einsatzbereiten „Fallback-Lösung“ kann ein für die Praxis nutzenstiftender Bestandteil von Forschungsarbeiten sein, die sich der vorliegenden Arbeit anschließen.

Weiterhin kann der Funktionsumfang eines WSBS im Rahmen einer Forschungsarbeit auf ein Mandanten-orientiertes System ausgeweitet werden. Unter einem Mandanten-orientierten WSBS wird ein IT-System verstanden, das für mehrere Unternehmen bzw. Organisationen das Erfassen, Bewahren und Bereitstellen einer WSB ermöglicht. Jede Organisation agiert dabei in einem eigenen und individuell für die Organisation verfügbaren Systembereich. Ein Mandant kann eigenständig Adressräume, Erfassungsbereiche und Kontext-Informationen definieren sowie Systemkonfigurationen vornehmen. Lediglich die Datenbasis des WSBS umfasst die Bestände aller Mandanten und bietet somit die Möglichkeit, Web-Site-Biographien mehrerer Unternehmen oder Organisationen zu kombinieren. Auf diese Weise lassen sich nicht nur einzelne Web Sites vollständig für

die Nachwelt konservieren, sondern auch die Lebensgeschichte von Web-Präsenzen kompletter Branchen oder Domänen erstellen.

Neben der Entwicklung eines Mandanten-fähigen WSBS kann die Verwendung der Blockchain-Technologie als Datenbasis für ein WSBS überprüft werden. Da sowohl bei klassischen Web-Archiven als auch bei einem WSBS die Integrität und Authentizität der Daten bzw. Abbilder eine wesentliche Rolle spielt, kann die Blockchain-Technologie ein geeignetes Fundament sein, um eine verlässliche und ausfallsichere Speicherung von Abbildern einzelner Web Sites zu ermöglichen. In diesem Kontext kann nicht nur die Frage nach der systematischen Entwicklung eines solchen Systems gestellt, sondern auch eine Bewertung des Nutzens von Blockchain-basierten System-Lösungen für die Langzeiterhaltung von organisationalen Web Sites vorgenommen werden.

*Für weiterführende Arbeiten, die sich auf den in der vorliegenden Arbeit skizzierten und umgesetzten Prototyp beziehen, wird der Quell-Code des Prototyps als Bestandteil der Dissertationsleistung mit eingereicht.*

## Literaturverzeichnis

1. **Abegglen, Martin:** Geschichtsforscher werten Archive mit maschinellem Lernen aus, Online im Internet: <https://www.heise.de/newsticker/meldung/Geschichtsforscher-werten-Archive-mit-maschinellm-Lernen-aus-3940801.html>, 29.01.2018.
2. **Adar, Eytan; Teevan, Jaime; Dumais, Susan T.; Elsas, Jonathan L.:** The web changes everything: understanding the dynamics of web content, in: Proceedings of the Second ACM International Conference on Web Search and Data Mining (WSDM 2009): Barcelona, Spain, February 9-12, 2009, S. 282-291.
3. **Ainsworth, Scott; AlSum, Ahmed; SalahEldeen, Hany; Weigle, Michele C.; Nelson, Michael L.:** How much of the Web is Archived?, Online im Internet: <http://arxiv.org/abs/1212.6177>, 15.05.2016.
4. **Akinici, Nick; Heidrich, Joerg:** Informationsverpflichtet, in: iX, 1/2018, S. 38-39.
5. **Alpar, Andre; Koczy, Markus; Metzen, Maik:** SEO - Strategie, Taktik und Technik: Online-Marketing mittels effektiver Suchmaschinenoptimierung, Wiesbaden: Springer Gabler 2015.
6. **Amarasinghe, Sean:** Service worker development cookbook: build highly available and performant native web applications that seamlessly integrate with third-party APIs, Packt Publishing Ltd 2016.
7. **Amazon.com, Inc (Hrsg.):** Amazon.de: Günstige Preise für Elektronik & Foto, Filme, Musik, Bücher, Games, Spielzeug & mehr, Online im Internet: <https://amazon.de>, 03.01.2020.
8. **Amtsblatt der Europäischen Union (Hrsg.):** VERORDNUNG (EU) 2016/679 DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG, Online im Internet: <http://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32016R0679&from=DE>, 05.02.2018.
9. **Aschenbrenner, Andreas; Rauber, Andreas:** Mining Web Collections, in: Web Archiving, Hrsg.: Masanès, Julien, Berlin, Heidelberg: Springer-Verlag 2010.
10. **AUDI AG (Hrsg.):** Audi e-tron, Online im Internet: <https://www.audi.de/de/brand/de/neuwagen/tron/audi-e-tron/linien-pakete.html>, 06.11.2019.
11. **Bächle, Michael; Lehmann, Frank R.:** E-Business: Grundlagen elektronischer Geschäftsprozesse im Web 2.0, München: Oldenbourg 2010.

12. **Balzert, Heide:** UML 2 in 5 Tagen: der schnelle Einstieg in die Objektorientierung, 2. Aufl., 1. Nachdruck, Herdecke: W3L-Verl 2010.
13. **Balzert, Heide:** Webdesign & Web-Ergonomie: Websites professionell gestalten, Herdecke: W3L 2004.
14. **Balzert, Helmut:** Lehrbuch der Software-Technik, 2. Aufl., 1. Nachdr., Heidelberg: Spektrum Akad. Verlag 2001.
15. **Balzert, Helmut:** Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements-Engineering, 3. Aufl, Heidelberg: Spektrum, Akad. Verl 2009.
16. **Bea, Franz Xaver; Göbel, Elisabeth:** Organisation: Theorie und Gestaltung, 4., neu bearbeitete und erweiterte Auflage, Stuttgart: Lucius & Lucius 2010.
17. **Becker, Jörg; Holten, Roland; Knackstedt, Ralf; Niehaves, Björn:** Forschungsmethodische Positionierung in der Wirtschaftsinformatik – epistemologische, ontologische und linguistische Leitfragen, in: Arbeitsberichte des Instituts für Wirtschaftsinformatik, Hrsg.: Institut für Wirtschaftsinformatik, Westfälische Wilhelms-Universität Münster, Arbeitsbericht Nr. 93, Münster: 2003.
18. **Beiner, Tobias; Hagenah, Ulrich; Kugler, Anna:** Es war einmal eine Website... - Kooperative Webarchivierung in der Praxis, in: o-bib, 1/2014, S. 291-304.
19. **Ben-David, Anat; Huurdeman, Hugo:** Web Archive Search as Research: Methodological and Theoretical Implications, in: Alexandria: The Journal of National and International Library and Information Issues, 25/2014, S. 93-111.
20. **Berger, Daniel:** Datenverlust: Myspace verliert riesiges Musikarchiv, Online im Internet: <https://www.heise.de/newsticker/meldung/Datenverlust-Myspace-verliert-riesiges-Musikarchiv-4338737.html>, 19.03.2019.
21. **Berghoff, Hartmut:** Moderne Unternehmensgeschichte: eine themen- und theorieorientierte Einführung, Paderborn: Schöningh 2004.
22. **Bergsmann, Johannes; Unterauer, Markus:** Requirements Engineering für die agile Softwareentwicklung: Methoden, Techniken und Strategien, 1. Aufl, Heidelberg: dpunkt-Verl 2014.
23. **Berners-Lee, Tim; Döpfner, Matthias:** Vage, aber HOCHINTERESSANT, in: WELT AM SONNTAG, 2017, S. 11-14.
24. **Bhasin, Shweta:** Web security basics, Cincinnati, Ohio: Premier Press 2003.
25. **Brown, Adrian:** Archiving websites: a practical guide for information management professionals, London: Facet Pub 2006.

26. **Brown, Adrian:** Practical digital preservation: a how-to guide for organizations of any size, London: Facet Publ 2013.
27. **Brügger, Niels:** Web History, an Emerging Field of Study, in: Web History, Hrsg.: Brügger, Niels, New York: Peter Lang Publishing 2010.
28. **Brügger, Niels:** Web History: An Analytical Grid, in: Web History, Hrsg.: Brügger, Niels, New York: Peter Lang Publishing 2010.
29. **Brügger, Niels:** When the present web is later the past: Web historiography, digital history, and Internet Studies, in: Historical Social Research/Historische Sozialforschung, 2012, S. 102-117.
30. **Brügger, Niels:** Web historiography and Internet Studies: Challenges and perspectives, in: new media & society, 15/2012, S. 752-764.
31. **Budras, Corinna:** Für immer verloren, in: Frankfurter Allgemeine Sonntagszeitung (F.A.S), 19/2018, S. 26.
32. **Bühler, Peter; Patrick Schlaich; Sinner, Dominik:** HTML5 und CSS3: Semantik - Design - Responsive Layouts, Berlin, Germany: Springer Vieweg 2017.
33. **Bühler, Peter; Schlaich, Patrick; Sinner, Dominik:** Webdesign: Interfacedesign - Screendesign - Mobiles Webdesign, Berlin: Springer Vieweg 2017.
34. **Bundesministerium der Justiz und für Verbraucherschutz (Hrsg.):** Gesetz über die Deutsche Nationalbibliothek (DNBG) § 3 Medienwerke, Online im Internet: [https://www.gesetze-im-internet.de/dnbg/\\_\\_3.html](https://www.gesetze-im-internet.de/dnbg/__3.html), 26.05.2017.
35. **Carlos, Gilberto:** Responsive Web Design with jQuery, Birmingham: Packt Publishing 2013.
36. **Chatfield, Tom; Freytag, Carl:** 50 Schlüsselideen digitale Kultur, Berlin: Springer Spektrum 2013.
37. **Cleve, Jürgen; Lämmel, Uwe:** Data mining, 2. Auflage, Berlin Boston: De Gruyter Oldenbourg 2016.
38. **Collins, Mark J.:** Pro HTML5 with CSS, JavaScript, and multimedia: complete website development and best practices, United States: Apress 2017.
39. **Consultative Committee for Space Data Systems (Hrsg.):** REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS), Online im Internet: <https://public.ccsds.org/pubs/650x0m2.pdf>, 12.08.2017.
40. **Copeland, Rick:** MongoDB applied design patterns, Sebastopol, CA: O'Reilly Media, Inc 2013.

41. **Crowls, Josh; Blank, Grant:** International hyperlinks in online news media, in: The web as history: using web archives to understand the past and the present, Hrsg.: Brügger, Niels; Schroeder, Ralph, London: UCL Press 2017.
42. **Cuy, Sebastian; Fischer, Martin; Peters, Jens; de Oliveira, Daniel; Puhl, Johanna; Rau, Lisa; Thaller Manfred (Hrsg.):** Das Digitale Archiv NRW in der Praxis: eine Softwarelösung zur digitalen Langzeitarchivierung, Hamburg: Kovač 2013.
43. **DAS EUROPÄISCHE PARLAMENT UND DER RAT DER EUROPÄISCHEN UNION (Hrsg.):** RICHTLINIE 2014/65/EU DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 15. Mai 2014 über Märkte für Finanzinstrumente sowie zur Änderung der Richtlinien 2002/92/EG und 2011/61/EU, Online im Internet: <http://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32014L0065&from=EN>, 01.12.2018.
44. **dejure.org Rechtsinformationssysteme GmbH (Hrsg.):** Rechtsprechung LAG Hessen, 24.01.2012 - 19 SaGa 1480/11, Online im Internet: <https://dejure.org/2012,733>, 03.02.2018.
45. **Dern, Gernot:** Management von IT-Architekturen: Leitlinien für die Ausrichtung, Planung und Gestaltung von Informationssystemen, 2., verb.erw. Aufl, Wiesbaden: Vieweg 2006.
46. **Deutscher Bundestag (Hrsg.):** Entwurf eines Gesetzes über die Deutsche Nationalbibliothek (DNBG), Online im Internet: <http://dipbt.bundestag.de/dip21/btd/16/003/1600322.pdf>, 05.05.2017.
47. **Doglio, Fernando:** REST API Development with Node.js: Manage and Understand the Full Capabilities of Successful REST Development, 2018.
48. **Ebert, Christof:** Systematisches Requirements Engineering: Anforderungen ermitteln, spezifizieren, analysieren und verwalten, 5., überarb. Aufl., Heidelberg: dpunkt-Verl 2014.
49. **eMarketer (Hrsg.):** eMarketer Updates Worldwide Internet and Mobile User Figures, Online im Internet: <https://www.emarketer.com/Article/eMarketer-Updates-Worldwide-Internet-Mobile-User-Figures/1015770>, 25.05.2017.
50. **Feliciati, Pierluigi; Dobрева, Milena (Hrsg.):** Access to digital archives: studying users' expectations and behaviours, in: Digital Archives: Management, use and access, London: Facet Publishing 2018.
51. **Ferstl, Otto K.; Sinz, Elmar J.:** Grundlagen der Wirtschaftsinformatik, 7., aktualisierte Aufl, München: Oldenbourg 2013.

52. **Fitch, K.:** Web site archiving: an approach to recording every materially different response produced by a Website, Sanctuary Cove, Australia: The Ninth Australian World Wide Web Conference 2003.
53. **Funk, Stefan E.:** Migration, in: nestor Handbuch: eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 2.3, 2010, Hrsg.: Neuroth, Heike; Oßwald, Achim; Scheffel, Regine; Strathmann, Stefan; Jehn, Mathias, 2., aktualisierte Auflage, Glückstadt: Verlag Werner Hülsbusch, Fachverlag für Medientechnik und -wirtschaft 2016.
54. **Gabriel, Roland; Weber, Peter; Schroer, Nadja; Lux, Thomas:** Basiswissen Wirtschaftsinformatik, 2. Aufl, Dortmund: W3L-Verl 2014.
55. **Gadatsch, Andreas:** Datenmodellierung für Einsteiger, Wiesbaden: Springer Fachmedien Wiesbaden 2017.
56. **Galitz, Wilbert O.:** The essential guide to user interface design: an introduction to GUI design principles and techniques, 3rd ed, Indianapolis, IN: Wiley Pub 2007.
57. **Gaus, Wilhelm:** Berufe im Archiv-, Bibliotheks-, Informations- und Dokumentationswesen: ein Wegweiser zur Ausbildung, Berlin: Springer 1986.
58. **Gesellschaft für Informatik e. V. (Hrsg.):** Digitales Kulturerbe, Online im Internet: <https://www.gi.de/themen/grand-challenges/digitale-kultur.html>, 28.03.2017.
59. **Google LLC (Hrsg.):** Building for VR on the Web, Online im Internet: <https://codelabs.developers.google.com/codelabs/webvr/index.html>, 19.04.2019.
60. **Goschler, Juliana:** Metaphern für das Gehirn: eine kognitiv-linguistische Untersuchung, Berlin: Frank & Timme 2008.
61. **Graham, Mark:** Robots.txt meant for search engines don't work well for web archives, Online im Internet: <https://blog.archive.org/2017/04/17/robots-txt-meant-for-search-engines-dont-work-well-for-web-archives/>, 05.05.2017.
62. **Grävemeyer, Arne:** Mit der Zeitmaschine in die Altstadt, in: c't magazin für computer und technik, 15/2019, S. 128–131.
63. **Große-Oetringhaus, Wigand F.:** Strategische Identität — Orientierung im Wandel, Berlin, Heidelberg: Springer Berlin Heidelberg 1996.
64. **Gruhn, Volker; Striemer, Rüdiger:** The essence of software engineering, Schweiz: Springer 2018.
65. **Haber, Peter:** Digital past: Geschichtswissenschaft im digitalen Zeitalter, München: Oldenbourg Verlag 2011.

66. **Hale, Scott A.; Blank, Grant; Alexander, Victoria D.:** Live versus archive: Co-mapping a web archive to a population of web pages, in: The web as history: using web archives to understand the past and the present, Hrsg.: Brügger, Niels; Schroeder, Ralph, London: UCL Press 2017.
67. **Hansch, Pierre; Rentschler, Christian:** EmotionWeb: emotionale Websites durch Bewegtbild und Sound-Design, Berlin: Springer Vieweg 2012.
68. **Harvey, D. R.; Gillian, Oliver:** Digital curation, Second edition, Chicago: ALA Neal-Schuman, an imprint of the American Library Association 2016.
69. **Hassler, Marco:** Web Analytics: Metriken auswerten, Besucherverhalten verstehen, Website optimieren, 3., aktualisierte und erw. Aufl, Heidelberg Hamburg: mitp, Verl.-Gruppe Hüthig, Jehle, Rehm 2012.
70. **Haumer, Florian; Donsbach, Wolfgang:** Der Wertschöpfungsbeitrag von Corporate Publishing: Effekte formaler und inhaltlicher Gestaltungsmerkmale von Kundenmagazinen, Wiesbaden: Springer VS 2013.
71. **Heidrich, Jörg; Heidrich, Joerg:** Pflichtverschiebung, in: iX, 1/2018, S. 44–46.
72. **Heinrich, Lutz J.; Heinzl, Armin; Roithmayr, Friedrich:** Wirtschaftsinformatik-Lexikon: mit etwa 4000 Stichwörtern und 3700 Verweistichwörtern, 7., vollständig überarbeitete und erw. Aufl., München: Oldenbourg 2004.
73. **Herbig, Daniel:** Google Stadia: Die Cloud-Gaming-Plattform, die Spieler über Youtube zocken lässt, Online im Internet: <https://www.heise.de/newsticker/meldung/Google-Stadia-Die-Cloud-Gaming-Plattform-die-Spieler-ueber-Youtube-zocken-laesst-4340325.html>, 10.04.2019.
74. **Hessischer Rundfunk (Hrsg.):** Gegen den Gedächtnisverlust: Das Archiv im digitalen Zeitalter, Online im Internet: <http://www.hr-inforadio.de/podcast/wissen/podcast-episode-5084.html>, 05.05.2017.
75. **Hillebrand, Ralf-Thomas:** Online-Kommunikation für Verbände: wie Ihre Botschaften die Zielgruppen sicher erreichen und überzeugen, Wiesbaden: Springer Gabler 2018.
76. **Hoffmann, Dirk W.:** Software-Qualität, Berlin: Springer 2008.
77. **International Internet Preservation Consortium:** The WARC Format 1.0, Online im Internet: <https://iipc.github.io/warc-specifications/specifications/warc-format/warc-1.0/#file-and-record-model>, 25.01.2020.

78. **Internet Archive (Hrsg.):** Internet Archive Wayback Machine, Online im Internet: [http://web.archive.org/web/20161119154605\\*/http://wiwi.uni-giessen.de/home/fb02/](http://web.archive.org/web/20161119154605*/http://wiwi.uni-giessen.de/home/fb02/), 28.03.2017.
79. **Internet Archive (Hrsg.):** Internet Archive Wayback Machine, Online im Internet: [https://web.archive.org/web/\\*/https://www.uni-giessen.de/fbz/fb02](https://web.archive.org/web/*/https://www.uni-giessen.de/fbz/fb02), 08.08.2020.
80. **Jochemich, Hanno:** „Fehler HTTP 404 – Seite nicht gefunden“ Oder: Zum Stand der Archivierung von Webinhalten in den öffentlich-rechtlichen Rundfunkanstalten, in: Das Gedächtnis des Rundfunks: die Archive der öffentlich-rechtlichen Sender und ihre Bedeutung für die Forschung, Hrsg.: Hasselbring, Bettina; Bernard, Birgit; Behmer, Markus, Wiesbaden: Springer VS 2014.
81. **Jones, Steve; Johnson, Camille:** Web Use and Web Studies, in: Web Archiving, Hrsg.: Masanès, Julien, Berlin, Heidelberg: Springer-Verlag 2010.
82. **jQuery Foundation (Hrsg.):** jQuery – write less, do more, Online im Internet: <https://jquery.com/>, 11.05.2018.
83. **Kappel, Gerti; Pröll, Birgit; Reich, Siegfried; Retschitzegger, Werner:** Web Engineering – Die Disziplin zur systematischen Entwicklung von Web-Anwendungen, in: Web engineering: systematische Entwicklung von Web-Anwendungen, Hrsg.: Kappel, Gerti, 1. Aufl, Heidelberg: dpunkt-Verl 2004.
84. **Kappes, Martin:** Netzwerk- und Datensicherheit: eine praktische Einführung, 2., überarb. und erw. Aufl, Wiesbaden: Springer Vieweg 2013.
85. **Keitel, Christian; Deutsches Institut für Normung (Hrsg.):** Vertrauenswürdige digitale Langzeitarchivierung nach DIN 31644, 1. Aufl, Berlin: Beuth 2013.
86. **Kilian, Thomas; Langner, Sascha:** Online-Kommunikation: Kunden zielsicher verführen und beeinflussen, 1. Aufl, Wiesbaden: Gabler 2010.
87. **Klahold, André:** Empfehlungssysteme: Recommender Systems – Grundlagen, Konzepte und Lösungen, 1. Aufl, Wiesbaden: Vieweg + Teubner 2009.
88. **Königs, Hans-Peter:** IT-Risikomanagement mit System: praxisorientiertes Management von Informationssicherheits-, IT- und Cyber-Risiken, 5., überarbeitete und erweiterte Auflage, Wiesbaden: Springer Vieweg 2017.
89. **Krause, Dirk:** VR im Web: Virtual-Reality-Web-Apps mit A-Frame programmieren, in: c't magazin für computer und technik, 16/2017, S. 116-122.
90. **Kreymer, Ilya:** Pywb Webrecorder web archive replay and capture tools, Online im Internet: <https://pypi.org/project/pywb/>, 13.12.2019.

91. **Krüger, Jörg Dennis; Kopp, Matthias:** Web-Content managen: professioneller Einsatz von Content-Management-Systemen, München: Markt-und-Technik-Verlag 2002.
92. **Liu, Alex X.; World Scientific (Firm):** Firewall design and analysis, Singapore; Hackensack, N.J.: World Scientific Pub. Co. 2011.
93. **Lobe, Adrian:** Wir sammeln eine Milliarde Seiten pro Woche, in: Frankfurter Allgemeine Zeitung (F.A.Z), 121/2015, S. 13.
94. **Ludewig, Jochen; Lichter, Horst:** Software Engineering: Grundlagen, Menschen, Prozesse, Techniken, 3., korrigierte Aufl., Heidelberg: dpunkt Verl. 2013.
95. **Masanès, Julien:** Web archiving, Berlin, London: Springer 2010.
96. **Masanès, Julien:** Web Archiving: Issues and Methods, in: Web Archiving, Hrsg.: Masanès, Julien, Berlin, Heidelberg: Springer-Verlag 2010.
97. **Meidl, Oliver:** Global Website: Webdesign im internationalen Umfeld, Wiesbaden: Springer Gabler 2013.
98. **Meier, Andreas:** Werkzeuge der digitalen Wirtschaft: Big Data, NoSQL & Co Eine Einführung in relationale und nicht-relationale Datenbanken, Wiesbaden: Springer Vieweg 2018.
99. **Meier, Andreas; Kaufmann, Michael:** SQL- & NoSQL-Datenbanken, 8., überarbeitete und erweiterte Auflage, Berlin Heidelberg: Springer Vieweg 2016.
100. **Meier, Andreas; Stormer, Henrik:** eBusiness & eCommerce: Management der digitalen Wertschöpfungskette, 3. Aufl, Berlin: Springer Gabler 2012.
101. **Meinel, Christoph; Sack, Harald:** WWW – Kommunikation, Internetworking, Web-Technologien, Berlin: Springer 2013.
102. **Melo, F.; Gomes, D.; Bicho, D.:** A Comparison Between The Performance of Wayback Machines, Online im Internet: <http://sobre.arquivo.pt/sobre/publicacoes-1/a-comparison-between-the-performance-of-wayback>, 13.12.2019.
103. **Menczer, Filippo:** Web Crawling, in: Web Data Mining – Exploring Hyperlinks, Contents, and Usage Data, Hrsg.: Liu, Bing, Berlin, New York: Springer-Verlag 2007.
104. **Mertens, Peter; Bodendorf, Freimut; König, Wolfgang; Schumann, Matthias; Hess, Thomas; Buxmann, Peter:** Grundzüge der Wirtschaftsinformatik, 12., grundlegend überarbeitete Auflage, Berlin: Springer Gabler 2017.
105. **Metzger, Franziska:** Geschichtsschreibung und Geschichtsdenken im 19. und 20. Jahrhundert, in: Historische Zeitschrift, Band 296, Heft 3/2013, S. 811.

106. **Meyer, Eric; Yasseri, Taha; Hale, Scott A.; Cowls, Josh; Schroeder, Ralph; Margetts, Helen:** Analysing the UK web domain and exploring 15 years of UK universities on the web, in: The web as history: using web archives to understand the past and the present, Hrsg.: Brügger, Niels; Schroeder, Ralph, London: UCL Press 2017.
107. **Miller, Steven J.:** Metadata for digital collections: a how-to-do-it manual, New York: Neal-Schuman Publishers 2011.
108. **Mitchell, Ryan E.:** Web scraping with Python: collecting data from the modern web, 2. Auflage, Beijing, Boston, Farnham, Sebastopol, Tokyo: O'Reilly Media 2018.
109. **Momigliano, Arnaldo:** The development of Greek biography, Expanded ed., 1st Harvard University Press pbk. ed, Cambridge, Mass: Harvard University Press 1993.
110. **MongoDB, Inc. (Hrsg.):** Relational Vs Non Relational Database, Online im Internet: <https://www.mongodb.com/scale/relational-vs-non-relational-database>, 22.04.2019.
111. **Mozilla and individual contributors (Hrsg.):** Headless mode, Online im Internet: [https://developer.mozilla.org/en-US/docs/Mozilla/Firefox/Headless\\_mode](https://developer.mozilla.org/en-US/docs/Mozilla/Firefox/Headless_mode), 25.01.2020.
112. **Nahari, Hadi; Krutz, Ronald L.:** Web commerce security: design and development, Indianapolis, Ind: Wiley 2011.
113. **Nentwich, Michael:** Cyberscience: research in the age of the internet, Vienna: Austrian Acad. of Sciences Press 2003.
114. **Noak, Pit:** Die Zukunft der Archive: Games nicht nur bewahren, sondern spielbar erhalten, Online im Internet: <https://www.heise.de/newsticker/meldung/Die-Zukunft-der-Archive-Games-nicht-nur-bewahren-sondern-spielbar-erhalten-4259331.html>, 29.12.2018.
115. **o. V.:** Blockchain – Meine Daten gehören mir, in: wisu - Das Wirtschaftsstudium, 4/2018, S. 344–346.
116. **Office of Government Commerce (OGC) (Hrsg.):** The official introduction to the ITIL service lifecycle, London: Stationary Office 2007.
117. **Vogel, Oliver (Hrsg.):** Software-Architektur: Grundlagen - Konzepte - Praxis, 2. Aufl, Heidelberg: Spektrum Akad. Verl 2009.

118. **Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften, DARIAH-DE (Hrsg.):** Handbuch Digital Humanities: Anwendungen, Forschungsdaten und Projekte / DARIAH-DE, Berlin: epubli 2015.
119. **Bengel, Günther; Baun, Christian; Kunze, Marcel; Stucky, Karl-Uwe (Hrsg.):** Masterkurs parallele und verteilte Systeme: Grundlagen und Programmierung von Multicore-Prozessoren, Multiprozessoren, Cluster, Grid und Cloud, 2. erw. und aktualisierte Auflage, Wiesbaden: Springer Vieweg 2015.
120. **Marshall, Brianna H. (Hrsg.):** The complete guide to personal digital archiving, Chicago: ALA Editions, an imprint of the American Library Association 2018.
121. **Oermann, Andrea; Jäschke, Gerald; Dittmann, Jana:** Vertrauenswürdige und abgesicherte Langzeitarchivierung multimedialer Inhalte, Frankfurt am Main: nestor - Kompetenznetzwerk Langzeitarchivierung c/o Deutsche Nationalbibliothek 2009.
122. **Pant, Gautam; Srinivasan, Padmini; Menczer, Filippo:** Crawling the Web, in: Web dynamics: adapting to change in content, size, topology and use ; with 29 tables, Hrsg.: Levene, Mark; Poulouvassilis, Alexandra, Berlin: Springer 2004.
123. **Pastukhov, Oleksandr; Dobрева, Milena (Hrsg.):** Legal issues surrounding digital archiving, in: Digital Archives: Management, use and access, London: Facet Publishing 2018.
124. **Patni, Sanjay:** Pro RESTful APIs: Design, Build and Integrate with REST, JSON, XML and JAX-RS, Berkeley, CA: Apress Imprint : Apress 2017.
125. **Pfeffers, Ken; Tuunanen, Tuure; Rothenberger, Marcus A.; Chatterjee, Samir:** A design science research methodology for information systems research, in: Journal of Management Information Systems, Vol. 24 Nr. 3/2007, S. 45-77.
126. **Picot, Arnold:** Richtungsdiskussionen in der Wirtschaftsinformatik, in: Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung, 62/2010, S. 662-679.
127. **Rathje, Ulf:** Technisches Konzept für die Datenarchivierung im Bundesarchiv, in: Der Archivar, 2/2002, S. 117-120.
128. **Recker, Jan:** Scientific research in information systems: a beginner's guide, Heidelberg: Springer 2013.
129. **Rogers, Richard; Brügger, Niels (Hrsg.); Milligan, Ian (Hrsg.):** Periodizing Web Archiving: Biographical, Event-Based, National and Autobiographical Traditions, in: The SAGE Handbook of Web History, Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne: SAGE Publishing 2018.

130. **Rosenthal, David; Robertson, Thomas; Lipkis, Tom; Reich, Vicky; Morabito, Seth:** Requirements for Digital Preservation Systems – A Bottom-Up Approach, Online im Internet: <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>, 12.08.2017.
131. **Rudolph, Simone:** Servicebasierte Planung und Steuerung der IT-Infrastruktur im Mittelstand: ein Modellansatz zur Struktur der IT-Leistungserbringung, Wiesbaden: Gabler 2009.
132. **Rumpel, Rainer J.:** Planung und Realisierung von IT-Infrastrukturen - ein prozessbasierter Ansatz, München: Oldenbourg 2012.
133. **Rupp, Chris:** Requirements-Engineering und -Management: professionelle, iterative Anforderungsanalyse für die Praxis, 5., aktualisierte und erw. Aufl, München: Hanser 2009.
134. **Sackmann, Steffan:** IT-Sicherheit, Online im Internet: <http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/technologien-methoden/Informatik--Grundlagen/IT-Sicherheit/index.html>, 01.03.2018.
135. **Sadalage, Pramod J.; Fowler, Martin:** NoSQL distilled: a brief guide to the emerging world of polyglot persistence, Upper Saddle River, NJ: Addison-Wesley 2013.
136. **Saleh, Iman:** Formalizing Data-Centric Web Services., Cham: Springer International Publishing 2015.
137. **Sammer, Petra:** Storytelling, 1. Aufl., 3. korrigierter Nachdruck, Beijing: O'Reilly 2015.
138. **Schicker, Edwin:** Datenbanken und SQL: eine praxisorientierte Einführung mit Anwendungen in Oracle, SQL Server und MySQL, 5., aktualisierte und erweiterte Auflage, Wiesbaden: Springer Vieweg 2017.
139. **Schiedermeier, Ute:** Wirtschaftsarchive am Beispiel des Siemens Historical Institute, in: AWW-Informationen, AWW-Informationen Special – Webarchivierung/2012, S. 12–13.
140. **Schneider, Holger:** Digitale Amnesie: Langzeitarchivierung digitaler Dokumente im betrieblichen Umfeld, Norderstedt: Books on Demand 2012.
141. **Schreyögg, Georg; Geiger, Daniel:** Organisation: Grundlagen moderner Organisationsgestaltung: mit Fallstudien, 6., vollständig überarbeitete und erweiterte Auflage, Wiesbaden: Springer Gabler 2016.

142. **Schrimpf, Sabine:** Das OAIS-Modell für die Langzeitarchivierung: Anwendung der ISO 14721 in Bibliotheken und Archiven, 1. Aufl, Berlin: Beuth 2014.
143. **Schug, Alexander:** History marketing: ein Leitfaden zum Umgang mit Geschichte in Unternehmen, 2., unveränd. Aufl., Bielefeld: transcript 2013.
144. **Schüle, Hubert:** Die Usability von Websites, in: wisu - Das Wirtschaftsstudium, 36/2007, S. 1092–1099.
145. **Schumann, Natascha:** Einführung in die digitale Langzeitarchivierung, in: Langzeitarchivierung von Forschungsdaten: Standards und disziplinspezifische Lösungen, Hrsg.: Althenhöner, Reinhard; Oellers, Claudia, Berlin: Scivero Verl. 2012.
146. **Schwenck, Rebecca:** Langzeitarchivierung digitaler Informationen: Ein internationaler Überblick über Organisation und Projektarbeit, Saarbrücken: VDM 2008.
147. **Schwickert, Axel C.:** Web Site Engineering: Ökonomische Analyse und Entwicklungssystematik für eBusiness-Präsenzen, Stuttgart: Teubner 2001.
148. **Sens, Bastian:** Schluss mit 08/15-Websites – so bringen Sie Ihr Online-Marketing auf Erfolgskurs: 79 Tipps für Ihren Online-Auftritt, Wiesbaden: Springer Gabler 2017.
149. **Springer Gabler Verlag (Hrsg.):** Stichwort: Website, Online im Internet: <http://wirtschaftslexikon.gabler.de/Archiv/75916/website-v12.html>, 11.07.2017.
150. **Statista (Hrsg.):** Number of internet users worldwide from 2005 to 2018 (in millions), 29.04.2019 Online im Internet: <https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/>, 29.04.2019.
151. **Statista (Hrsg.):** PYPL. Die beliebtesten Programmiersprachen weltweit laut PYPL-Index im Januar 2018., Online im Internet: <https://de.statista.com/statistik/daten/studie/678732/umfrage/beliebteste-programmiersprachen-weltweit-laut-pypl-index/>, 01.02.2018.
152. **Steinke, Tobias:** Herausforderungen bei der Langzeitverfügbarkeit von Webarchiven, in: AWV-Informationen, AWV-Informationen Special – Webarchivierung/2012, S. 34–37.
153. **Steyer, Manfred; Softic, Vildan:** AngularJS: moderne Webanwendungen und Single Page Applications mit JavaScript, Beijing Cambridge Farnham Köln Sebastopol Tokyo: O'Reilly 2015.
154. **Stubblebine, Tony:** Regular expression pocket reference, 2. Ausgabe, Sebastopol, Calif: O'Reilly 2007.

155. **The Apache Software Foundation (Hrsg.):** CouchDB, Online im Internet: <https://couchdb.apache.org/>, 13.05.2020.
156. **Theobald, Elkre; Häusler, Annika:** Die Website als zentrales Element der digitalen Markenführung, in: Brand Evolution: moderne Markenführung im digitalen Zeitalter, Hrsg.: Haisch, Philipp T., Wiesbaden: Gabler 2011.
157. **Thesmann, Stephan:** Interface design: usability, user experience und accessibility im Web gestalten, 2., aktualisierte und erweiterte Auflage, Wiesbaden: Springer Vieweg 2016.
158. **Ullrich, Dagmar:** Bitstream Preservation, in: nestor Handbuch: eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 2.3, 2010, Hrsg.: Neuroth, Heike; Oßwald, Achim; Scheffel, Regine; Strathmann, Stefan; Jehn, Mathias, 2., aktualisierte Auflage, Glückstadt: Verlag Werner Hülsbusch, Fachverlag für Medientechnik und -wirtschaft 2016.
159. **UNESCO (Hrsg.):** Charta zur Bewahrung des digitalen Kulturerbes, Online im Internet: <http://www.unesco.de/infothek/dokumente/unesco-erklaerungen/charta-zur-bewahrung-des-digitalen-kulturerbes.html>, 26.05.2017.
160. **Vogel, Oliver; Arnold, Ingo; Chughtai, Arif; Ihler, Edmund; Kehrer, Timo; Mehlig, Uwe; Zdun, Uwe:** Software-Architektur: Grundlagen - Konzepte - Praxis, 2. Aufl, Heidelberg: Spektrum Akad. Verl 2009.
161. **von Zimmermann, Christian:** Geschichte oder Biographie: Leopold Rankes Porträts Papst Pauls III. und Wallensteins, in: Die Biographie – Beiträge zu ihrer Geschichte, Hrsg.: von Hemecker, Wilhelm, Berlin, New York: De Gruyter 2009.
162. **W3C - World Wide Web Consortium (Hrsg.):** The principles of unobtrusive JavaScript, Online im Internet: [https://www.w3.org/wiki/The\\_principles\\_of\\_unobtrusive\\_JavaScript](https://www.w3.org/wiki/The_principles_of_unobtrusive_JavaScript), 03.02.2018.
163. **W3C - World Wide Web Consortium (Hrsg.):** WHAT IS CSS?, Online im Internet: <https://www.w3.org/Style/CSS/>, 10.02.2018.
164. **Weill, Peter; Broadbent, Marianne:** Leveraging the new infrastructure: how market leaders capitalize on information technology, Boston, Mass: Harvard Business School Press 1998.
165. **Weitekamp, Lea; T3N (Hrsg.):** JavaScript, Flash, CSS und mehr: Die Geschichte des Webdesigns in 9 GIFs, Online im Internet: <https://t3n.de/news/javascript-flat-geschichte-webdesign-585903/>, 16.02.2018.

166. **Wirtz, Bernd W.; Becker, Daniel R.:** Geschäftsmodellansätze und Geschäftsmodellvarianten im Electronic Business, in: WiSt – Wirtschaftsstudium, Heft 2/2002, S. 85-90.
167. **Wöhr, Heiko:** Web-Technologien: Konzepte - Programmiermodelle - Architekturen, 1. Aufl, Heidelberg: dpunkt-Verl 2004.
168. **Xavier, Roche:** Copying Websites, in: Web Archiving, Hrsg.: Masanès, Julien, Berlin, Heidelberg: Springer-Verlag 2010.
169. **Ziehl, Stefan:** Langzeitarchivierung im digitalen Zeitalter Speichermedien, Strategien und Ausblicke, Saarbrücken: AV Akademikerverlag 2012.
170. **Zierau, Eld; Peyrard, Sébastien:** Digital Preservation Metadata in a Metadata Ecosystem, in: Digital preservation metadata for practitioners, New York, NY: Springer Berlin Heidelberg 2016.
171. **Zschau, O.; Traub, D.; Zahradka, R.:** Web Content Management. Websites professionell planen und betreiben. 2., überarb. und erw. Aufl, Bonn: Galileo Press 2002.
172. **Zschunke, Peter; Schmidt, Julia:** Digitalisierung: Museen sind bereit zu teilen, Online im Internet: <https://www.heise.de/newsticker/meldung/Digitalisierung-Museen-sind-bereit-zu-teilen-3609717.html>, 28.03.2017.