

Vorwort

Digitale Bibliotheken sind aus unserer global vernetzten Welt nicht mehr wegzudenken. Virtuelle Bibliotheksbesuche, spontane Zugriffe auf internationale Dokumentenbestände und verteilte Suchen über Verbundkataloge sind nur einige Beispiele für Aktivitäten, die wir ganz bequem zu Hause oder am Arbeitsplatz ausüben. Die technischen Entwicklungen auf diesem Gebiet sind rasant, die in den Verbunden nachgewiesenen Dokumentenbestände wachsen explosionsartig und die von Forschern angedachten zukünftigen Dienstleistungen sind atemberaubend.

*Nutzen digitaler
Bibliotheken*

Bei aller Freude über die zukünftigen Dienstleistungen bergen rasante technische Entwicklungen und explosionsartiges Wachstum aber große Risiken. Dies gilt in besonderem Maße, wenn digitale Dokumente über einen längeren Zeitraum in gleicher Art und Weise zugreifbar, in ihrem Erscheinungsbild nachweisbar authentisch und als digitale Artefakte der Nachwelt erhalten bleiben sollen. Mit den damit zusammenhängenden Fragen bei der Langzeitarchivierung beschäftigt sich dieses Buch. In den vereinigten Staaten wird dieser Problematik schon große Aufmerksamkeit geschenkt, vgl. etwa (Bergeron 2002).

*Probleme bei der
Langzeitarchivierung*

Diese unserer Meinung nach so wichtigen und drängenden Fragen bei der Langzeitarchivierung elektronischer Dokumente erfordern Kommunikation und Zusammenarbeit zwischen Personen unterschiedlichster Berufsgruppen: Politiker, Juristen, Wirtschaftswissenschaftler und vor allem natürlich Bibliothekare, Archivare und Informatiker müssen gemeinsam an einer Lösung arbeiten. Wir stellen zunächst das organisatorische Umfeld und verschiedene technische Lösungsansätze allgemeinverständlich dar. Dadurch soll eine gemeinsame begriffliche Basis für die Diskussion geschaffen werden. Später nähern wir uns den mit der technischen Umsetzung befassten Personen und beschreiben detaillierter einzelne Ansätze und aktuelle Entwicklungen.

»Stakeholders«

Das Buch wendet sich folgerichtig an eine breite Leserschaft aus unterschiedlichen Disziplinen, denen die Sorge und die Suche

*Zielgruppe und
Leserschaft*

nach Sicherung und Bewahrung des gemeinsamen digitalen Kulturguts gemeinsam sind. Zu nennen wären hier etwa

- ❑ Bibliothekare, Archivare und Historiker zur Schärfung des Problembewusstseins,
- ❑ Politiker, Juristen nebst Wirtschaftswissenschaftlern zur Benennung auf Maßnahmen zur Verbesserung des organisatorischen Umfelds und nicht zuletzt
- ❑ Informatiker und andere an den »neuen Medien« Forschende und Entwickelnde zur Bereitstellung der notwendigen technischen Infrastruktur.
- ❑ Studierende technischer Disziplinen profitieren beim Lesen dieses Buches von den detaillierten Beschreibungen einschlägiger Projekte.
- ❑ Studierende geisteswissenschaftlicher Fächer erweitern ihr Problemverständnis durch den möglichen Brückenschlag zu den eher technischen Problemen der Langzeitarchivierung.
- ❑ Wir wollen auch nicht vergessen all diejenigen aufzurütteln, denen die Bewahrung digitaler Daten nicht nur als Kulturgut am Herzen liegt, sondern auch aus rein ökonomischen Überlegungen heraus. Manager in der Industrie, in Banken und Versicherungen müssen einen langfristigen Zugriff auf ihre geschäftsrelevanten Daten verantworten und sind deshalb für viele Aspekte, die in diesem Buch angesprochen werden, wichtige Adressaten.
- ❑ Nach Lektüre des Buches sind nach Meinung der Autoren alle interessierten Leser in die Lage versetzt, die Notwendigkeit *jetzt* zu handeln zu spüren, den aktuellen Forschungsstand und die noch zu klärenden Fragenstellungen zumindest zu erahnen sowie realistische Vorgehensweisen zu erkennen.
- ❑ Den Experten wird mit diesem Buch ein umfassendes Kompendium zur Langzeitarchivierung geliefert.

Aufbau des Buches

Das Buch gliedert sich in zwei Teile. Der erste Teil umfasst sechs Kapitel, in denen wir einen methodischen Überblick geben.

Kapitel 1

Kapitel 1 führt in die Problematik der Langzeitarchivierung digitaler Dokumente ein. Wir diskutieren Probleme der Haltbarkeit elektronischer Speichermedien ebenso wie die zahlreichen Abhängigkeiten digitaler Dokumente von spezifischer Interpretationssoftware. Die wichtigsten technischen Ansätze zur Lösung der Probleme werden einführend vorgestellt. Juristische und gesellschaftliche Fragen, die auf dem Weg zu einer tragfähigen Lösung

der gesamten Archivierungsproblematik digitaler Dokumente beantwortet werden müssen, beschließen dieses einführende Kapitel.

Ein Referenzmodell, das zu einem internationalen Standard für Langzeitarchivierungssysteme führen könnte, wird mit dem so genannten Open Archival Information System Reference Model (OAIS) vorgestellt. Kapitel 2 schafft damit die Grundlage für eine klare und präzise Begrifflichkeit, wenn es gilt, konkrete Lösungsansätze zu bewerten und den unterschiedlichen Erhaltungsstrategien zuzuordnen.

Kapitel 2

Kapitel 3 erläutert die Migration als Erhaltungsstrategie und stellt die unterschiedlichen Verfahrensvarianten in der Umsetzung bei der Langzeitarchivierung vor. Diese Varianten umfassen Verfahren zur Auffrischung digitaler Medien ebenso wie die Migration auf andere digitale und nicht-digitale Medien. Wir kategorisieren digitale Dokumente und geben einen kurzen Überblick über archivierungstaugliche Standardformate.

Kapitel 3

Kapitel 4 erläutert die Emulation als die zweite wichtige Erhaltungsstrategie. Diese Strategie zielt darauf ab, die Originalabspielumgebungen digitaler Dokumente durch Emulation auf künftigen Computern dauerhaft am Leben zu erhalten. Wir erörtern verschiedene Varianten an Beispielen und diskutieren danach die Chancen und Risiken des Emulationsansatzes.

Kapitel 4

In Kapitel 5 gehen wir näher auf die Beschreibung und Auszeichnung von Dokumenten (engl. *Markup*) ein. Nach einer kurzen Klassifikation unterschiedlicher Markup-Varianten stellen wir Anwendungsmöglichkeiten im Bereich der Langzeitarchivierung vor. Abschließend wenden wir uns dem Problem der Persistenz digitaler Daten zu, das sich immer dann ergibt, wenn auf Daten verwiesen wird, die sich mit der Zeit verändern.

Kapitel 5

Standardisierte Dokumentenbeschreibungs-Sprachen für die Auszeichnung digitaler Dokumente mit Metadaten sind Gegenstand der Diskussionen in Kapitel 6. Zunächst behandeln wir die Möglichkeiten zur eingebetteten strukturellen Dokumentenbeschreibung, wobei wir genauer auf die Markup-Sprache XML eingehen werden. Danach zeigen wir einige Möglichkeiten der semantischen Dokumentenbeschreibung auf und erörtern damit verbundene technische Schwierigkeiten. Weiterhin stellen wir mit dem so genannten *Semantic Web* eine attraktive und viel diskutierte Erweiterung des heutigen World Wide Web vor.

Kapitel 6

Kapitel 7 beschließt den ersten Teil des Buches mit einer Diskussion der vorgestellten Methoden. Wir stellen einen kombinier-

Kapitel 7

ten Ansatz zur Langzeitarchivierung digitaler Dokumente vor, der uns aus heutiger Sicht vielversprechend erscheint.

Der zweite Teil widmet sich in drei Kapiteln einschlägigen nationalen wie internationalen Projekten. Sie illustrieren beispielhaft diejenigen Techniken, die im ersten Teil tutoriell angesprochen wurden. Die Wahl der Projekte erfolgte dabei nach reinen Überdeckungsgesichtspunkten und erhebt daher keinen Anspruch auf Vollständigkeit. Der interessierte Leser findet eine umfassende aktuelle Übersicht über fast alle laufenden einschlägigen Projekte unter <http://www.nla.gov.au/padi/>.¹

Kapitel 8 In Kapitel 8 stellen wir verschiedene Initiativen und Projekte vor, die sich mit der Auszeichnung von Dokumenten in der Praxis beschäftigen. So wird mit Pandora beispielsweise ein Projekt beschrieben, das sich ganz explizit dem Speichern und Archivieren von Hyperdokumentmengen im World Wide Web widmet.

Kapitel 9 Kapitel 9 beleuchtet die Migration anhand vier konkreter Projekte: Wir besprechen hierbei die Victorian Electronic Records Strategy (VERS-Projekt) ebenso wie die beiden Projekte »Preserving the Whole« und »Risk Management of Digital Information«. Letzteres sieht als das brennendste Problem nicht die Instabilität von Medien oder überalterte Hardware an, sondern die Veraltung von Datenformaten und zugehöriger Software. Schließlich beschreiben wir ein umfangreiches Projekt zur Problematik der Datenbankmigration. Erste Ergebnisse einer Umsetzbarkeitsuntersuchung werden vorgestellt.

Kapitel 10 Kapitel 10 beschreibt zunächst ein Emulationsexperiment, wie es im Rahmen des europäischen Verbundprojekts NEDLIB an der Nationalbibliothek der Niederlande (Koninklijke Bibliotheek) durchgeführt wurde. Anschließend stellen wir einen Ansatz auf Basis des so genannten Universal Virtual Computer (UVC) vor. Der UVC-Ansatz gibt eine erste Antwort auf die Risiken des Migrationsansatzes und auf die Bedenken, die gegen den ursprünglichen Emulationsansatz vorgebracht werden.

Danksagung Wir bedanken uns bei der Leitenden Direktorin der Universitätsbibliothek der Universität der Bundeswehr in München, Frau Dr. Stefanie Seidel, die uns in steter Bereitschaft kompetent und freundlich unterstützte. Ihr haben wir es auch zu verdanken, dass Herr Thomas Pieruschka Teile des Abschnitts über OAIS beisteuern konnte. Hierfür gebührt ihm unser Dank. Wir bedanken

¹Preserving Access to Digital Information (PADI).

uns ferner freundschaftlich bei Frau Dr. Marianne Dörr,² Leiterin der Hessischen Landesbibliothek Wiesbaden, die uns mehrere Jahre als Projektpartnerin begleitete und immer mit Sachverstand, Ausdauer und Charme für Diskussionen zur Verfügung stand.

Wir bedanken uns bei unserem ehemaligen Mitarbeiter Knut Weißer für seine Vorarbeiten zum Thema Migration und bei Dr. Wolfram Kahl für seine Beiträge und Hinweise zu Dokumentenbeschreibungs-Sprachen.

Die Deutsche Forschungsgemeinschaft (DFG) unterstützte ein Projekt zum Thema *Entwicklung eines organisatorisch-technischen Konzepts für die Langzeitarchivierung digitaler Publikationen in Bibliotheken*, das wir zusammen mit der Bayerischen Staatsbibliothek als Projektpartnerin bearbeiten durften. Ohne diese initiale Projektförderung wären die Ergebnisse des Buches weniger substanziell und in der Breitenwirkung weniger attraktiv ausgefallen. So wurden beispielsweise Projektergebnisse in einem internationalen Workshop diskutiert und Seminare abgehalten. Im Frühjahr 2000 sowie im Winter 2002 fanden an der Universität der Bundeswehr München einschlägige Seminare zu diesem Thema statt. Wir danken den Teilnehmern Dominik Christian Bódi, Thomas Domschke, Jens Ferchland, Matthias Frank, Sebastian Hess, Thomas Kiesling, Alexander Koenen, Stephan Moritz, Danny Poppe, Mirko Pracht, Daniel Prenzel, Patrick Roiss, Daniel Volk sowie Michael Wellhöfer für ihre Vorträge und eingereichten Ausarbeitungen, die eine erste Grobgliederung für dieses Buch erlaubten.³

München, im Juni 2003

U.M.B., P.R.
J.S., L.S.

²ehemalige Leiterin Digitalisierung und VD17 an der Bayerischen Staatsbibliothek.

³Der geneigte Leser findet unter <http://www2.informatik.unibw-muenchen.de/Events/Seminare/> sowie <http://www2.informatik.unibw-muenchen.de/LZA/index.html> detailliertere Informationen.