

**Besondere Herausforderungen bei der retrospektiven  
Digitalisierung historischen Pressematerials am Beispiel eines  
Projekts des Hamburgischen Welt-Wirtschafts-Archivs**

vorgelegt von

Kirsten Jeude, Kurs C/2004

im thematischen Schwerpunkt Informations-Management

am 20.02.2005

## Abstract

---

Die Retrodigitalisierung von Archivmaterialien und Bibliotheksbeständen für die Literatur- und Informationsversorgung hat in den letzten Jahren an Bedeutung gewonnen. Aufgrund der unterschiedlichen Medien und Gegenstände gibt es keine allgemein gültigen Lösungen für die Digitalisierung von Archivgut. Am Beispiel des Projekts zur retrospektiven Digitalisierung der Zeitungsausschnitte der Pressearchive des Hamburgischen Weltwirtschafts-Archivs (HWWA) wird gezeigt, welche besonderen Strategien und Lösungen möglich und nötig sind, um eine Presseauschnittssammlung in digitaler Form im Internet bereitstellen zu können.

Anhand von Projektberichten weiterer, von der Deutschen Forschungsgemeinschaft im Rahmen des Programms „Retrospektive Digitalisierung von Bibliotheksbeständen“ geförderter Digitalisierungsprojekte werden die wesentlichen Probleme und möglichen Lösungsstrategien aufgezeigt, die sich im Verlauf der Produktion und Bereitstellung elektronischer Daten auf der Grundlage historischer Druckschriften, Druckgraphiken und Handschriften einstellen können. Es werden einerseits bisher gewählte Lösungsansätze des HWWA-Projekts zu den Erfahrungen jener Projekte in Bezug gesetzt und andererseits werden mögliche Strategien für zukünftige Aufgaben des Projekts erarbeitet.

Herangezogen werden Kriterien wie Gegenstand, Medium, Materialmenge und Quelleigenschaften der einzelnen Projekte und die jeweilig gewählten Problemlösungsansätze. Neben dem Aspekt der Auswahl des Materials, die von der jeweiligen Motivation zur Konversion geleitet wird, besteht der Prozess der Digitalisierung im Wesentlichen aus drei Teilprozessen: Erfassen, Erschließen und Bereitstellen der Dokumente. Da für die Bereitstellung historischer, aber nicht urheberrechtsfreier Dokumente (hier: Zeitungsartikel und Aufsätze in Zeitschriften aus dem 20. Jh.) besondere Maßnahmen ergriffen werden müssen, wird zudem der Umgang mit dem Urheberrecht in die Untersuchung einbezogen.

## Inhaltsverzeichnis

---

1	Einleitung.....	1
1.1	Fragestellung .....	1
1.2	Methodik .....	1
1.3	Begriffsbestimmungen .....	3
2	Das DFG-Projekt zur retrospektiven Digitalisierung historischer Presseartikel auf Papier, Rollfilm und Mikrofiches der Archive des HWWA.....	4
2.1	Projektbeschreibung .....	4
2.2	Das Teilprojekt zur retrospektiven Digitalisierung des Personenarchivs .....	5
2.3	DFG-Förderprogramm „Retrospektive Digitalisierung von Bibliotheksbeständen“ ..	6
3	Aspekte der Retrodigitalisierung historischen Materials.....	8
3.1	Auswahl des Materials – Motivation für die Retrodigitalisierung .....	8
3.1.1	Wissenschaftliche Relevanz .....	8
3.1.2	Zugänglichmachung benutzungsbeschränkter Bestände .....	8
3.1.3	Erzeugung informationeller Mehrwerte .....	9
3.1.4	Bestandserhaltung und Langzeitarchivierung .....	9
3.1.5	Motivation für die Auswahl der historischen Pressearchive.....	10
3.2	Erfassen: Der technische Prozess der Digitalisierung - Scannen und Speichern.....	12
3.2.1	Arten der Digitalisierung .....	12
3.2.2	Speicherung der Scann-Ergebnisse .....	13
3.2.3	Strukturdaten.....	14
3.2.4	Scannen der Presseartikel bei der Digitalisierung des Personenarchivs.....	14
3.3	Inhaltliche Erschließung der Digitalisate.....	16
3.3.1	Erschließungskonzepte .....	16
3.3.2	Für und Wider Volltextdigitalisierung .....	19
3.3.3	Erschließen der digitalisierten Presseauschnitte des Personenarchivs .....	21
3.4	Bereitstellung und Präsentation der digitalisierten Dokumente .....	25
3.4.1	Bereitstellung: technische Infrastruktur und inhaltliche Aspekte.....	25
3.4.2	Präsentation des digitalisierten Personenarchivs.....	27
3.5	Exkurs: Urheberrechtsproblematik .....	28
3.5.1	Retrodigitalisierungsprojekte und das Urheberrecht.....	28
3.5.2	Die Urheberrechtsproblematik bei der Digitalisierung des Personenarchivs ...	30
4	Rückblick und Ausblick: Die besonderen Herausforderungen bei der Digitalisierung des historischen Personenarchivs des HWWA .....	32
5	Literatur .....	35
6	Anhang .....	37

## **1 Einleitung**

Seit Mitte der 1990er Jahre revolutioniert die Digitalisierung von Information die Informationsverbreitung und beeinflusst damit in starkem Maße die Aufgaben der Bibliotheken und Archive:

„Digitalisierung und genuin digitale Information bedeuten einen Qualitätssprung in der Informationsversorgung für Wissenschaft und Wirtschaft, sie verändern die Informationsverbreitung in ähnlicher Weise, wie die Erfindung des Buchdrucks die damalige Welt verändert hat [...] Digitalisierung verbessert die Nutzungsmöglichkeiten, insbesondere von schwer zugänglichen Informationen, wie Bildern, Handschriften oder seltenen Drucken. Digitalisierung von Informationen und genuin digitalen Informationen steigern durch ständige Verfügbarkeit und durch komfortable und effiziente Rechensoftware mit einheitlicher Oberfläche den schnellen Zugriff auf die volle Information, auf detaillierte Metadaten und relevante Verlinkung“ (Hänel 2000:131).

Die Bedeutung der Digitalisierung von Archiv- und Bibliotheksbeständen für die Literatur- und Informationsversorgung hat die Deutsche Forschungsgemeinschaft bereits 1997 bewogen, solche Projekte zu fördern, die mit der Digitalisierung historischer Bestände und deren Bereitstellung im Internet eine verbesserte Nutzung verfolgen.

### **1.1 Fragestellung**

Ausgehend von der Frage, inwieweit man die bisherigen Erfahrungen anderer Digitalisierungsprojekte heranziehen kann, um Strategien für ein seit Juni 2004 laufendes Retrodigitalisierungsprojekt am Hamburgischen Welt-Wirtschafts-Archiv (HWWA) zu entwickeln, zeigt sich, dass allgemeine Problemlösungsansätze für diesen Bereich nicht vorhanden sind:

„Jedes Digitalisierungsprojekt stellt neue und individuelle Anforderungen, so dass die einzelnen Arbeitsschritte jeweils genau aufeinander abgestimmt werden müssen.“

„Digitalisierungsprojekte haben eine vorlagenabhängige Eigendynamik, die jeweils individuell beschrieben und umgesetzt werden muß“ (Brantl 2003:14; 16).

In dieser Arbeit werde ich daher die besonderen Herausforderungen des am HWWA angesiedelten DFG-Projekts zur „Retrospektiven Digitalisierung der historischen Pressearchive des HWWA“ untersuchen. Da eine Auseinandersetzung mit den vielfältigen Aufgaben des Gesamtprojekts den Rahmen dieser Arbeit sprengen würde, werde ich mich hauptsächlich auf das Teilprojekt der Digitalisierung des Personenarchivs beziehen und nur punktuell auf das Gesamtprojekt eingehen.

### **1.2 Methodik**

Ich werde die bis zum jetzigen Stand des Projekts gewählten Lösungsansätze zu den Erfahrungen anderer Projekte des Förderprogramms zur „Retrospektiven Digitalisierung

von Bibliotheksbeständen“ in bezug setzen und darüber hinaus für zukünftige Aufgaben innerhalb des Projekts anhand jener Erfahrungen Problemlösungsstrategien entwickeln. Dazu werde ich beispielhaft Literatur über Retrodigitalisierungsprojekte sowie Internetangebote von solchen Projekten aus dem Pool von ca. 80 im Rahmen o.g. DFG-Programms geförderten Projekten zur „Retrospektiven Digitalisierung von Bibliotheksbeständen“ heranziehen und auch internationale Digitalisierungsprojekte einbeziehen.

Ich werde Kriterien wie Gegenstand, Medium, Materialmenge und Quelleneigenschaften der einzelnen Projekte zu ihren Problemlösungsansätzen in den Bereichen Erfassen, Erschließen und Bereitstellung in bezug setzen.

Exemplarisch werde ich die wesentlichen Probleme und möglichen Lösungsstrategien aufzeigen, die im Verlauf der Produktion und Bereitstellung elektronischer Daten auf der Grundlage historischen Druckschriften (und Druckgraphiken), aber auch Handschriften erscheinen können.

Im Rahmen dieser Einleitung werde ich im Folgenden zentrale Begriffe definieren. Im zweiten Kapitel werde ich die Rahmenbedingungen des Projekts zur Digitalisierung des historischen Personenarchivs beschreiben: Nach einer kurzen Vorstellung des Förderprogramms der Deutschen Forschungsgemeinschaft (DFG) werde ich Gegenstand und Ziele des Gesamtprojekts zur retrospektiven Digitalisierung der historischen Pressearchive des HWWA vorstellen. In einem kurzen Überblick werde ich das Teilprojekt zur Digitalisierung des Personenarchivs beschreiben und den derzeitigen Stand der Projektarbeit nennen.

Im dritten Kapitel werde ich die Aspekte der Digitalisierung historischen Materials untersuchen: Nach dem Aspekt der Auswahl des Materials, die von der jeweiligen Motivation zur Digitalisierung geleitet wird, sind im Wesentlichen drei Komponenten zu betrachten: Die Erfassung, die Erschließung und die Bereitstellung des Materials. Diese drei Komponenten sind inhaltlich und technisch sehr eng miteinander vernetzt, so dass eine trennscharfe Abgrenzung nicht in Gänze möglich ist.

In einem Exkurs werde ich die Bedeutung des Urheberrechts für die Bereitstellung historischer, aber nicht urheberrechtsfreier Dokumente für Retrodigitalisierungsprojekte untersuchen.

Im letzten Kapitel werde ich zusammenfassend die besonderen Herausforderungen des Projekts zur retrospektiven Digitalisierung des historischen Personenarchivs des HWWA benennen und einen Ausblick auf kommende Aufgaben, wie auch Probleme und mögliche Lösungsansätze geben.

### 1.3 Begriffsbestimmungen

Der Begriff „Digitalisierung“ wird je nach Kontext als rein technischer Prozess der Transformation analoger Informationen in digitale Informationen verstanden:

„Digitalisierung (...) meint zunächst, dass Texte in maschinenlesbare Form überführt werden, so dass sie mittels EDV benutzbar sind.“ (Ritzi 2000:61)

„Digitalisierung bezeichnet zweifelsfrei eine Aktion, an deren Ende nach einem Umformungsprozess ein digitaler (...) Informationsträger vorliegt, der ein Dokument enthält, das ursprünglich analog gespeichert war und nun als Sekundärform mit gleichem Inhalt existiert.“ (Leskien 2000:133)

Darüber hinausgehend wird „Digitalisierung“ als ein komplexer Arbeitsprozess verstanden, an dessen Ende ein Produkt dokumentarischer Arbeit entsteht: nämlich ein formal wie inhaltlich erschlossenes Dokument, das digital weltweit bereitgestellt werden kann:

„Dabei sind unter Digitalisierung im GDZ [Göttinger Digitalisierungszentrum] alle Vorgänge zu verstehen, die das analoge, gedruckte Buch in digitaler Form zum Benutzer bringen. D.h. Digitalisierung bedeutet nicht nur das Erzeugen von Images, sondern umfasst auch alle Vorgänge der Strukturierung und Erstellung von Metadaten sowie die Bereitstellung der digitalisierten Literatur im Internet“ (Enders 2001:225).

Letztere Definition beinhaltet neben dem Einbezug dokumentarischer Arbeit auch ein besonderes Kennzeichen digitalisierter Medien: Sie können leichter als ihre Druckoriginale und ohne Informations- und Qualitätsverluste weltweit verbreitet werden.

Der Begriff „Retrospektiv“ bezieht sich bei diesem Prozess auf das Alter der digitalisierten Dokumente:

„Retrospektiv meint in diesem Zusammenhang, dass die Publikation des Originalmediums bereits zurückliegt und dass die Sekundärform nicht gleichzeitig bzw. absichtsvoll parallel erscheint“ (Leskien 2000:133).

In dieser Arbeit wird der rein technische Transformationsprozess und die Erzeugung von bildhaften Darstellungen des Originaldokuments als „Scannen“ bezeichnet. Der Begriff „Digitalisierung“ wird im folgenden als Gesamtprozess verstanden, der neben dem technischen Vorgang des Scannens, also dem Erfassen, auch aus der Bestandsauswahl, dem Erschließen und der Bereitstellung der ausgewählten Dokumente besteht.

„Retrospektive Digitalisierung“ soll im Folgenden als Prozess der Auswahl, des digitalen Erfassens, Erschließens und Bereitstellens historischer Quellen verstanden werden.

## 2 Das DFG-Projekt zur retrospektiven Digitalisierung historischer Presseartikel auf Papier, Rollfilm und Mikrofiches der Archive des HWWA

### 2.1 Projektbeschreibung

Im Gesamtprojekt „Retrospektive Digitalisierung von historischen Presseartikeln auf Papier, Rollfilm und Mikrofiche der Archive des HWWA“ sollen die Albestände der Pressearchive (ca. 1880 bis 1930) digitalisiert und in einer webfähigen Datenbank angeboten werden.

Es handelt sich um einen umfangreichen, historisch wertvollen und teilweise einmaligen Bestand von Presseartikeln aus ca. 700 Zeitungen und Zeitschriften. Der Grundstein dieser Archive wurden im Hamburgischen Kolonialinstitut gelegt. Das Hamburgische Welt-Wirtschafts-Archiv wurde am 20.10.1908 als Zentralstelle des Kolonialinstituts in Hamburg gegründet. In der frühen Zeit der Archive wurden überwiegend Artikel über die deutschen Kolonien und ihre Verwaltung, koloniale Waren und Firmen gesammelt. Der Bestand bildet so die deutsche Kolonialgeschichte im Spiegel der Presse ab. Nach und nach erweiterte sich das Spektrum der Themen und Quellen: Die Presseartikel umfassen nun ökonomische und politische Themen, die Bereiche Wissenschaft, Forschung, Bildung und Kultur aus weltweiter Sicht.

Einerseits soll das Material vor dem Verfall gerettet werden, andererseits sollen die Bestände, die derzeit nur vor Ort in den Archiven des HWWA eingesehen werden können, der Öffentlichkeit möglichst komfortabel zugänglich gemacht werden.

Das Material liegt in Form von Papier und auf Rollfilm vor und unterliegt dabei hoher Abnutzung. Insbesondere die Benutzung der entsprechenden Geräte zum Lesen der Rollfilme ist kompliziert und nicht nutzerfreundlich. Zudem ist der inhaltliche Zugriff auf die alten Dokumente ohne die kompetente Lektoratsberatung nicht möglich, da die alten Systematiken bekannt sein und interpretiert werden müssen.

Die Pressearchive des HWWA sind in folgende Teilbestände unterteilt:

1. Das **Sach-, Länder- und Warenarchiv** enthält Presseartikel über ökonomische und politische Themen sowie die Bereiche Wissenschaft, Forschung, Bildung, Kultur und Soziales bezogen auf einzelne Regionen, Nationen, Ländergruppen, Kontinenten und die Welt sowie Presseartikel über Produkte, Märkte, Rohstoffe, Halbfertig- und Fertigerzeugnisse bezogen auf einzelne Länder, Kontinente und die Welt.
2. Das **Firmenarchiv** enthält Presseartikel und Geschäftsberichte über ca. 8.000 internationale Firmen.

Insgesamt handelt es sich bei diesen beiden Beständen für den Zeitraum bis 1930 um 2,7-2,9 Mio. Dokumente, die auf etwa 2750 Rollfilmen vorliegen.

3. Das **Personenarchiv** besteht aus ca. 120.000 Original-Presseausschnitten zu – für den Zeitraum bis 1930 – ca. 3.200 bedeutenden Persönlichkeiten.

Diese drei verteilt gespeicherten Sammlungen von Presseinformationen sollen als ein gemeinsames Informationsangebot bereitgestellt werden. Die web-basierte Datenbank soll gut strukturierte und inhaltlich erschlossene Pressedaten zur Verfügung stellen, so dass die thematische Anfrage in Eigenregie durchgeführt werden kann.

Das Gesamtprojekt zerfällt in verschiedene Teilprojekte, die sich zum Einen aus den einzelnen Sammlungen ergeben. Zum Anderen erfordern die unterschiedlichen Medien, auf denen das Material vorliegt, getrennte bzw. unterschiedliche Arbeitsprozesse: Firmen- sowie Sach-, Länder- und Warenarchiv wurden bereits im Zuge von Bestandserhaltsmaßnahmen in den 1960er Jahren verfilmt. Die Originale liegen heute nicht mehr vor, da sie im Anschluss an die jeweilige Sicherheitsverfilmung aus Platzmangel vernichtet wurden. Hier werden die vorliegenden Mikrofilme digitalisiert. Das Personenarchiv liegt ausschließlich auf Papier vor, so dass eine Direktdigitalisierung vom Original möglich ist.

## **2.2 Das Teilprojekt zur retrospektive Digitalisierung des Personenarchivs**

Im Personenarchiv des HWWA werden seit etwa 1900 Zeitungs- und Zeitschriftenauschnitte über Personen der Zeitgeschichte (Politiker, Wirtschaftsgrößen, Staatsoberhäupter, Militär, Wissenschaftler, Künstler etc.) gesammelt. Vereinzelt liegen Artikel aus dem 19. Jh. vor, die sich überwiegend mit Kolonialbeamten befassen.

Das Archiv weist für den Zeitraum bis 1930 ca. 120.000 Dokumente auf, die für die Digitalisierung in Frage kommen.

Die Artikel sind in verschiedensten Frakturlettern, z.T. auch in Antiqua – je nach Quelle und Jahr – gesetzt. Die Artikel liegen auch in nichteuropäischen Schriften (z.B. kyrilisch) vor und sind außer in deutscher auch in englischer, französischer, spanischer, portugiesischer, italienischer, finnischer, niederländischer, russischer Sprache und weiteren Sprachen verfasst. Teilweise enthalten sie Fotografien. Das Druckbild ändert sich von Ausgabe zu Ausgabe und von Quelle zu Quelle.

Die Artikel werden chronologisch geordnet in alphabetisch nach Personennamen sortierten Mappen aufbewahrt. Es handelt sich um originale Presseauschnitte, die auf Träger-



papier (Größe etwa DIN A4) aufgeklebt worden sind. Vereinzelt finden sich auch Flugblätter, Plakate, Denkschriften, Portraitaufnahmen und ganze Zeitschriftenhefte.

Die Dokumente sind von sehr heterogenem Umfang und in disparatem Zustand: Teilweise entspricht eine Seite einem Artikel, teilweise gibt es Folgeseiten. Das Papier ist dann doppelseitig beklebt oder erstreckt sich über bis zu 4 DIN A4-Seiten, die am Rand zusammengeheftet sind. Mitunter sind Artikel, deren Größe über A4 hinausgehen, gefaltet. Teilweise liegen gesamte Zeitungsseiten vor, deren Größe über A3 hinausgeht. Das Material ist mit Flüssigkleber auf dem Trägerpapier fixiert, teils mit mehr oder minder transparentem Folienpapier an das Papier geheftet. Oft finden sich handschriftliche Vermerke. Durch äußere Einflüsse sind die Vorlagen oft wellig, eingerissen oder in sonst einer Weise fragil. Teilweise sind die Lettern durch Feuchtigkeitseinwirkung infolge der Sturmflut von 1962 verschwommen.

Die Besonderheit des Personenarchivs – die Informationen liegen auf dem Medium Papier vor – macht eine Bestandserhaltungsmaßnahme des vom Zerfall bedrohten Papiers notwendig. Darüber hinaus können die Inhalte des Personenarchivs im Zuge des Digitalisierungsprozesses mit den Beständen der weiteren Archive verknüpft werden.

Zum gegenwärtigen Zeitpunkt sind Maßnahmen zur Vorbereitung des Materials für das Scannen bei einem Dienstleister für ca. vier Fünftel des Bestandes durchgeführt. Etwa zwei Drittel der Dokumente sind bereits gescannt. Eine Maske zur formalen und inhaltlichen Erschließung des Materials befindet sich in der Entwicklung und ein erster Entwurf einer Webseite für die Präsentation der Images liegt vor.

### **2.3 Das DFG-Förderprogramm „Retrospektive Digitalisierung von Bibliotheksbeständen“**

Das o.g. Gesamtprojekt wird im Bereich Programm und Infrastrukturförderung, Wissenschaftliche Literaturversorgung und Informationssysteme 3 (LIS 3) von der Deutschen Forschungsgemeinschaft (DFG) zunächst für 2 Jahre gefördert.

Die Deutsche Forschungsgemeinschaft begann bereits 1997 Förderprogramme für die retrospektive Digitalisierung auszuloben. Sie sind als Teil der angestrebten „Verteilten Digitalen Forschungsbibliothek“<sup>1</sup> aufzufassen.

---

<sup>1</sup> Die „verteilte digitale Forschungsbibliothek“ ist ein Förderbereich der Deutschen Forschungsgemeinschaft, unter der die DFG ihre verschiedenen Programme zum Einsatz der elektronischen Datenverarbeitung und der Digitalisierung zusammengeführt hat. Gedruckte Medien waren nur in einem arbeitsteiligen Verfahren, z.B. im verteilten System der überregionalen Literaturversorgung der DFG, in ihrer Gesamtheit einigermaßen vollständig zu sammeln und bereitzustellen. Digitale Medien erleichtern Sammlung und Bereitstellung zwar dadurch, dass sie ohne Informations- und Qualitätsverlust weltweit verbreitet

Es wurden bereits im ersten Jahr 3 Mio. DM bereitgestellt, da man erkannte, dass die digitalisierte Bereitstellung älterer Literatur eine wesentliche Verbesserung ihrer Zugänglichkeit wie ihrer Nutzbarkeit für konkrete Forschungsaufgaben bedeutet (vgl. Mittler 1997:83).

Retrospektive Digitalisierung soll jedoch nicht nur die Benutzung erleichtern, sondern auch den gezielten Zugriff verbessern:

„Die Aufbereitung und Präsentation sollen (...) auch einen definierten „Mehrwert“ in Zugriff und Benutzung gegenüber den Printversionen erzielen“ (Dörr 2000:154).

Zusammenfassend wird der Gegenstand des Förderprogramms „Retrospektive Digitalisierung von Bibliotheksbeständen“ folgendermaßen definiert:

„Gegenstand des Förderprogramms ist die retrospektive Digitalisierung ausgewählter Bibliotheksbestände zur Ermöglichung eines direkten Zugriffes auf einen elektronischen Gesamtbestand. Ziel des Programms ist die Verbesserung der Literaturversorgung durch den Einsatz digitaler Technik. Im Vordergrund stehen dabei der Direktzugriff auf für Forschung und Lehre wichtige Bestände, der Mehrfachzugriff auf vielgenutzte Literatur, die digitale Bereitstellung schwer zugänglicher Bestände und die erweiterte Nutzung bisher nur wenig bekannter Materialien“ ([http://gauss.suub.uni-bremen.de/html/dfg\\_foerderung.html](http://gauss.suub.uni-bremen.de/html/dfg_foerderung.html); 31.01.05).

Im Rahmen dieses Förderprogramms wurden in Göttingen an der Staats- und Universitätsbibliothek und in München an der Bayerischen Staatsbibliothek Service- und Kompetenzzentren aufgebaut, welche die Retrodigitalisierungsprojekte unterstützen. Komplexität und Tempo der technischen Entwicklung und die Notwendigkeit der Einführung von Standards ließen dies sinnvoll erscheinen.

Die Bestände der Projekte, die unter diesem Förderprogramm gefördert werden, sind sehr vielfältig und liegen auf unterschiedlichsten Medien vor: Es handelt sich in den meisten Fällen um Druckschriften und Handschriften:

„Als Objekte einer retrospektiven Digitalisierung kommen sowohl ältere gedruckte oder handschriftliche Quellenbestände wie auch neuere oder aktuelle Publikationen in Frage“ (Erber 1997:143).

Gefördert werden aber auch Projekte der Digitalisierung von Bildern<sup>2</sup>, Karten<sup>3</sup>, Objekten (Ostraka<sup>4</sup>, Papyri<sup>5</sup>, Druckplatten<sup>6</sup>), Tonaufnahmen<sup>7</sup> oder bewegten Bildern<sup>8</sup> (Film und Video).

---

werden können, doch wegen der außerordentlich steigenden Zahl digital verbreiteter wissenschaftlicher Texte und Medien sind aber die verteilte Sammlung und Bereitstellung sowie auch die verteilte Archivierung von besonderer Bedeutung (vgl. Mittler 1997:81).

<sup>2</sup> Z.B. das Projekt „Bildarchiv Foto Marburg d. Philipps-Universität / Deutsches Dokumentationszentrum für Kunstgeschichte“ des Nationales Bildarchiv zur Kunst und Architektur in Deutschland; <http://www.bildindex.de/>

Ebenso werden Projekte zur Erarbeitung von Arbeitsprozessen<sup>9</sup> und solche zur digitalen inhaltlichen Erschließung<sup>10</sup> im Rahmen dieses Programms gefördert (vgl. Anhang 1).

### **3 Aspekte der Retrodigitalisierung historischen Materials**

Vor dem gewöhnlich dreigeteilten Vorgehen im Prozess der Digitalisierung, Erfassen, d.h. Scannen (vgl. Kap. 3.2), strukturelle und inhaltliche Erschließung (vgl. Kap. 3.3.) und Bereitstellung der Dokumente (vgl. Kap. 3.4) muss von der antragstellenden Institution zunächst eine Entscheidung über die zu digitalisierenden Bestände getroffen werden.

#### **3.1 Auswahl des Materials – Motivation für die Retrodigitalisierung**

##### **3.1.1 Wissenschaftliche Relevanz**

Die Auswahl des Materials geschieht zunächst aufgrund grundsätzlicher Überlegungen in Bezug auf die Bedeutung der historischen Bestände für die Forschung. Über die bisherigen Zugangsmöglichkeiten hinaus soll ein Direktzugriff auf die Inhalte forschungsrelevanter Bestände erreicht werden. Wie bisherige Erfahrungen zeigen, gibt es allerdings von Fachgebiet zu Fachgebiet höchst unterschiedliche Bewertungen dessen, was relevant ist, so dass eine universal gültige Festlegung für retrospektive Prioritätensetzung nicht möglich ist (vgl. Leskien 2000:135). Dies traf allerdings bereits in ähnlicher Weise für die Auswahl analoger Materialien zu.

##### **3.1.2 Zugänglichmachung benutzungsbeschränkter Bestände**

Eine weiterer Aspekt ist die Zugänglichkeit benutzungsbeschränkter Bestände. Bisher schwer zugängliche Bestände sollen allgemein zugänglich gemacht werden, damit z.B.

---

<sup>3</sup> z.B. „Retrospektive Digitalisierung des historischen Kartenbestandes der SuUB Bremen“; <http://gauss.suub.uni-bremen.de>

<sup>4</sup> Ostraka = Tonscherben, die in der Antike als Schreibmaterial gebraucht wurden; z.B. das Projekt „Erfassung, Erschließung, Digitalisierung und Bereitstellung der Gießener Ostraka“ an der Universitätsbibliothek Gießen; <http://digibib.ub.uni-giessen.de/cgi-bin/populo/ost.pl>

<sup>5</sup> z.B. das Projekt „Digitalisierung der Kölner Papyrusbestände“ am Institut für Altertumskunde der Universität zu Köln; <http://www.uni-koeln.de/phil-fak/ifa/NRWakademie/papyrologie/index.html>

<sup>6</sup> z.B. das Projekt „Erfassung von Braille-Musik-Matrizen in elektronisch lesbarer Form“ des Verein zur Förderung der Blindenbildung e.V., Hannover; [www.vzfb.de](http://www.vzfb.de)

<sup>7</sup> z.B. das Projekt „Digitale Archivierung von Tonaufnahmen semitischer Sprachen und Dialekte und deren Publizierung im Internet“ des Seminars für Sprachen und Kulturen des Vorderen Orients - Semitistik - der Universität Heidelberg; <http://semitistik.uni-hd.de/tondoku.htm>

<sup>8</sup> z.B. das Projekt „Bewegtbildkommunikation im Internet“ des Bibliotheks- und Informationssystem der Universität Oldenburg; <http://www.bis.uni-oldenburg.de/Projekte/dvd/index.html>

<sup>9</sup> z.B. das Projekt „Workflow und Werkzeuge zur Digitalen Bereitstellung größerer Mengen von Archivgut“ der Landesarchivdirektion Baden-Württemberg; [http://www.landearchiv-bw.de/sixcms/detail.php?template=hp\\_artikel&id=4972&id2=4583&sprache=de](http://www.landearchiv-bw.de/sixcms/detail.php?template=hp_artikel&id=4972&id2=4583&sprache=de)

wenig bekannte Materialien eine größere Aufmerksamkeit finden und einer erweiterten Nutzung zugeführt werden können. Hier sind vor allem Dokumente zu nennen, deren Originale besonders schützenswert sind (Handschriften, Inkunabeln, seltene Drucke, historische Landkarten usw.). Ihre digitale Bereitstellung und Präsentation kann auch neue Qualitäten schaffen. Ein Beispiel hierfür ist das Projekt zur Digitalisierung der Giessener Papyrus- und Ostrakasammlung, bei der neben der weltweiten Zugänglichkeit auch eine tiefere Erschließung und Verknüpfung mit Volltexten erricht wurde (vgl. <http://www.uni-giessen.de/ub/ueber/papyri.html#beschreibung>; 16.02.05).

### **3.1.3 Erzeugung informationeller Mehrwerte**

Insbesondere der Aspekt der Schaffung eines informationellen Mehrwerts dient als Kriterium zur Auswahl des Materials. Im Rahmen von Digitalisierungsprojekten ist es selbstverständlich geworden, sich nicht darauf zu beschränken lediglich die originalen Informationen vom analogen auf einen digitalen Informationsträger zu kopieren und in dieser Form zugänglich zu machen. Dieses Vorgehen ist unwirtschaftlich und wird als Aufwand gewertet, der die eigentliche Nachfrage nicht trifft. Digitalisate bleiben hierbei zumeist sogar hinter den Eigenschaften der Originale zurück (vgl. Leskien 2000:139). Vielmehr sind durch die elektronische Speicherung Mehrwerte zu schaffen, z.B. bei Texten durch die Möglichkeit zu einer Volltextrecherche. Häufig gehen erzeugte informationelle Mehrwerte jedoch auch Hand in Hand mit einer tieferen Erschließung, die eventuell auch in konventioneller Form möglich gewesen wäre, jedoch nie den Komfort hätten erreichen können, der durch die digitale Verarbeitung und Speicherung möglich wird.

### **3.1.4 Bestandserhaltung und Langzeitarchivierung**

Eine letzte Motivation für die Digitalisierung ist die Informationssicherung und Langzeitarchivierung. Viele Digitalisierungsprojekte, welche die Digitalisierung wertvoller, seltener oder einmaliger Bestände wie Handschriftensammlungen, Papyrussammlungen o.ä. zum Gegenstand haben, nennen als eines ihrer Ziele die Sicherung und den Schutz der Dokumente.

Die Verfilmung vom Verfall bedrohter Dokumente galt bisher als „Königsweg“ zur Bestandserhaltung. Die Konversion auf Film birgt jedoch Nachteile: Die Nutzung mittels Lesegerät ist mühsam und belastend für die Augen. Die Preise für Filmduplikate

---

<sup>10</sup> z.B. das Projekt „Festkultur Online“ der Herzog-August-Bibliothek Wolfenbüttel; <http://www.hab.de/forschung/projekte/festkultur.htm>

und Lesegeräte sind so hoch, dass keine private Anschaffung möglich ist. Folglich muss sich der Nutzer auf lange Wege zur Bibliothek, begrenzte Öffnungszeiten, mögliche Wartezeiten und eine begrenzte Anzahl von Lesegeräten einstellen. Ein Ziel – die Bestandserhaltung – ist mit der Verfilmung zu erreichen, an der schlechten Verfügbarkeit der Bestände hat sich dadurch allerdings nichts geändert:

„Die Verfilmung von Bibliotheksmaterialien hat somit zwar zum Erhalt seltener Bibliotheksbestände beigetragen, aber die Verbesserung der Literaturversorgung hielt sich in Grenzen“ (Ritzi 2000, S:62).

Im Gegensatz dazu ermöglicht Digitalisierung (und Bereitstellung im Internet) zwar eine weltweite Verfügbarkeit, die jedoch nach wie vor ungeklärte Frage, wie die dauerhafte Erhaltung der digitalen Dokumente gesichert werden kann, lässt an der Tauglichkeit der Digitalisate als Medien der Langzeitarchivierung zweifeln. Bisher stehen als Informationsträger mit Langzeitcharakter nach wie vor vor allem Papier und Mikrofilm im Vordergrund.

Es muss bei der Entscheidung für eine Digitalisierung als Mittel der Langzeitarchivierung also eine Abwägung der verschiedenen Zielsetzungen vorgenommen werden.

Neben den bisher genannten Überlegungen zur wissenschaftlichen Relevanz, verbesserter Zugänglichkeit, Erzeugung von informationellem Mehrwert und Bestandserhaltung sollten auch unternehmerische Gesichtspunkte nicht aus den Augen verloren werden, denn:

„Digitalisierungsprojekte sind eine Form des electronic publishing, Bibliotheken werden also verlegerisch tätig. Die Auswahl eines für die Nutzer attraktiven und von der Durchführung her bewältigbaren Bestands ist eine ganz wesentliche Voraussetzungen für den Erfolg“ (Dörr 2000:163).

### **3.1.5 Motivation für die Auswahl der historischen Pressearchive**

Mit der Aufarbeitung der Presseartikel der Jahre 1880 bis 1930 der Archive des HWWA soll eine Entwicklungsepoche in wirtschaftsgeschichtlicher Perspektive dargestellt werden:

„Die im Pressearchiv des HWWA gesammelten und inhaltlich zugeordneten Artikel aus [rund 700] internationalen Zeitungen und Zeitschriften zeigen als Originalquellen den Wandel der Wirtschaftssysteme und Wirtschaftsformen über das gesamte vergangene Jahrhundert. (...) (Sie) dokumentieren wirtschaftlich relevante Themen, Ereignisse, Unternehmen und Personen des internationalen Zeitgeschehen im Spektrum von politischer Willensbildung und vielfältiger Meinungsäußerung. Dabei werden die von der Wirtschaft geprägten Sozialmilieus besonders deutlich“ (aus dem Projektantrag: Nordmann; Striefler 2003:5).

Hauptziele des Projekts sind die formale und inhaltliche Erschließung der Presseartikel und ihre recherchefähige Präsentation im Internet, so dass diese noch weithin unbekannt Dokumente weltweit zugänglich gemacht werden können.

Wissenschaft, Forschung und Wirtschaft sind auf den schnellen, vollständigen, komfortablen und kostensparenden Zugang zu Informationen angewiesen. Da Zeitungsverlage nur kurzfristig vollständigen oder auszugsweisen Zugriff auf ihre Produkte ermöglichen und Hosts oft nur ein spezielles und begrenztes Angebot an Verlagsinformationen kostenpflichtig zur Verfügung stellen, kann das HWWA dagegen sachlich und chronologisch geordnete Dokumente seit ca. 1880 zu einem breiten Themenspektrum anbieten, die zukünftig mittels einer einheitlichen Rechercheoberfläche (Suchmaske) recherchiert werden können.

Die getrennt gesammelten Bestände können durch die Digitalisierung zusammengeführt werden. Dies ermöglicht eine bisher nicht leicht mögliche bestandsübergreifende Recherche, womit ein informationeller Mehrwert geschaffen wird. Dazu sollen auch bestandsübergreifende thematische Dokumentationsprofile entwickelt werden. Durch die Bereitstellung des gut strukturierten und inhaltlich erschlossenen Materials in einer web-basierten Datenbank kann jederzeit und vielerorts darauf zugegriffen werden. Damit unterstützt dieses Angebot Forschungsaufgaben, Lehre und Studium.

„Die Presseoriginaltexte sind prädestiniert als Wissensbasis für wissenschaftliche Einrichtungen, sie dienen der zeitgeschichtlichen Bildung, und Unternehmen und Organisationen steht ein hochwertiges Archiv zur Erarbeitung z.B. von Festschriften und geschichtlichen Abrissen zur Verfügung“ (aus dem Projektantrag: Nordmann; Striefler, 2003:12).

Das Angebot richtet sich damit primär an wissenschaftlich orientierte Nutzer, die zu Forschungszwecken auf den Bestand zugreifen, sekundär stünde das Material der interessierten (Fach-)Öffentlichkeit zur Verfügung.

Ein letztes Ziel des Digitalisierungsprojektes ist die Bestandserhaltung: Wie in Kap. 2.1 und 2.2 bereits erwähnt, sind die Papiervorlagen des Personenarchivs vom Zerfall bedroht und auch die Sicherheitsverfilmungen des Firmen- sowie Sach-, Länder- und Warenarchivs zeigen bereits Verfallserscheinungen, so dass vermehrt Master-Filme ausgeliehen werden müssen. Durch die durch langjährige Lagerung verursachten Farbveränderungen der Filme ist die Lesetauglichkeit bereits stark herabgesetzt. Zudem entsprechen die Lesegeräte nicht den heutigen Anforderungen, die an eine bequeme und schnelle Nutzung gestellt werden. Zur Bewahrung dieses nahezu lückenlosen Bestandes ist eine Konversion notwendig. In Abwägung der Nutzenstiftung einer Digitalisierung

gegenüber einer bloßen Verfilmung wurde die Digitalisierung als Methode der Bestandssicherung gewählt.

Es wurde das Material bis 1930 ausgewählt, weil hier eine kritische Grenze sowohl im Hinblick auf die Bestandssicherung als auch – aus urheberrechtlichen Gründen – im Hinblick auf die elektronische Präsentation der Dokumente gegeben ist.

Zusammenfassend kann man sagen, dass mit der Digitalisierung der historischen Pressarchive des HWWA folgendes erreicht werden soll:

- Der vom Zerfall bedrohte Bestand kann erhalten werden;
- die Bearbeitung neuer wissenschaftlicher Fragestellungen wird möglich;
- auf das bisher schwer zugängliche Material kann unabhängig von Zeit und Raum zugegriffen werden und
- bisher kaum bekanntes Material kann der Öffentlichkeit zugänglich gemacht werden.

### **3.2 Erfassen: Der technische Prozess der Digitalisierung - Scannen und Speichern**

#### **3.2.1 Arten der Digitalisierung**

Die Arbeitsgruppe „Technik“ des Förderprogramms „Retrospektive Digitalisierung von Bibliotheksbeständen“ hat Richtlinien als Arbeitshilfe für Digitalisierungsprojekte erarbeitet (vgl. Mittler 1998:27f.):

Die einfachste Form und kostengünstigste Methode der Digitalisierung von Dokumenten ist das seitenweise Scannen der Dokumente, wodurch eine Folge von Stillbildern (Images) entsteht. Hierbei müssen Entscheidungen bezüglich des Grads der Auflösung (zwischen 150 und 600 dpi<sup>11</sup>) sowie über Bitonalität, Graustufen oder Farbe getroffen werden<sup>12</sup>.

Daneben steht die weit kostintensivere Volltextdigitalisierung mit Hilfe von OCR<sup>13</sup> oder die manuelle Texterfassung (vgl. Leskien 2000:144f.).

Denkbar ist auch der Weg über den Mikrofilm und die anschließende Digitalisierung des Mikrofilms. Beim diesem Verfahren kann der Film gleichzeitig als Sicherungsme-

---

<sup>11</sup> dpi = dots per inch

<sup>12</sup> Bei der Wahl der Auflösung sollte darauf geachtet werden, ob eine spätere Texterkennung mittels OCR gewünscht ist. Hier kann eine hochwertige Auflösung eher zu schlechteren Ergebnissen führen (vgl. Erber 1997:152).

<sup>13</sup> Optical Character Recognition = Optische Texterkennung

dium dienen und das Digitalisat als Benutzungsmedium, zudem ist die Digitalisierung vom Film kostengünstiger als das direkte Scannen vom Original. Auf der Basis eines einmal angefertigten Films ist bei verbesserter Digitalisierungstechnik jederzeit eine erneute Umsetzung des Films ohne Rückgriff auf das Original möglich (vgl. Erber 1997:150).

Die Ergebnisse der Entscheidungen für eine Art der Digitalisierung beeinflussen sowohl die Kosten wie auch das benötigte Speicherplatzvolumen und die Ladezeiten. Letztere sind ein wichtiges Kriterium für die Nutzerzufriedenheit.

Kriterien für die Wahl der Digitalisierungsmethode sind die Art der Vorlage (Wert, Alter, Qualität/Erhaltungszustand) sowie das gewünschte Resultat (Qualität der Wiedergabe am Bildschirm/im Ausdruck, erforderliche Auflösung, Wiedergabe in Schwarz-Weiss, Graustufen oder Farbe) (vgl. Erber 1997:150).

### **3.2.2 Speicherung der Scann-Ergebnisse**

Für die Speicherung der Images haben sich einige De-facto-Standards etabliert, die von vielen Digitalisierungsprojekten übernommen wurden (vgl. u.a. Dörr 2000; Effinger et.al. 2003; Klinge; Schüler 2003): Das TIFF<sup>14</sup>-Format zur Archivierung und als Vorlage für die Umwandlung in Präsentationsformate für das Internet wie JPG<sup>15</sup> und GIF<sup>16</sup>. Das TIFF-Format erlaubt eine originalgetreue Reproduktion des Originals.

Für Textdaten empfiehlt sich eine Speicherung in einem plattformunabhängigen Format wie ASCII oder das UTF-8–Unicode-Format. Letzteres spielt aufgrund der multilingualen Zeichenunterstützung und in Verbindung mit XML<sup>17</sup> eine zentrale Rolle (vgl. Brantl 2003:15). XML ist als plattformunabhängiges Format auf kein besonderes Betriebssystem angewiesen und ist daher besonders für den Datenaustausch zwischen verschiedenen Betriebssystemen geeignet, wie auch für die Langzeitarchivierung digitaler Daten. Für die Präsentation von Text bzw. Text und Bild im Internet können zudem Dateien im PDF<sup>18</sup>-Format angeboten werden.

---

<sup>14</sup> TIFF = Tagged Image File Format = Zusammen mit Encapsulated Postscript ist es das wichtigste Format zum Austausch von Daten in der Druckvorstufe

<sup>15</sup> JPG/JPEG = Joint Photographic Experts Group = verlustbehaftete Kompressions-Verfahren für digitale Bilder

<sup>16</sup> GIF = Graphics Interchange Format = digitales Bildformat mit guter verlustfreier Komprimierung für Bilder mit geringer Farbtiefe (2 bis 256 Farben); Standardformat für Pixeldateien

<sup>17</sup> XML = Extended Markup Language = Auszeichnungssprache bzw. -grammatik

<sup>18</sup> PDF = Portable Document Format = gibt die Dokumente des Ursprungsprogramms einschließlich aller Schriften, Farben, Grafiken und Bildern präzise wieder



Als Trägermedium bieten sich Massenspeicher und Compact-Disks (CD-ROM<sup>19</sup> und DVD<sup>20</sup>) an, wobei die Eignung digitaler Speicher als Langzeitarchivierungsmedien noch nicht als abschließend geklärt gelten kann (vgl. Mittler 1997:83). Ein einigermaßen zukunftssicheres Speichermedium ist derzeit der magneto-optische Speicher, dessen Haltbarkeit mit 30 Jahren beziffert wird (vgl. <http://www.speicherguide.de/magazin/optical.asp?theID=705>; 17.02.2005).

### 3.2.3 Strukturdaten

Bereits bei der Vorbereitung der Dokumente für den Prozess des Scannens muss an die strukturierte Aufbereitung der Images gedacht werden. Die den Images zugeteilten Dateinamen können die durch Drucker und Buchbinder vorgegebene Anordnung der Seiten ersetzen. Da jedoch in den meisten Fällen mehr als eine Quelle im Rahmen eines Projektes digitalisiert wird, müssen neben der Binnenstruktur des Dokuments zusätzliche Strukturmerkmale in den Dateinamen aufgenommen werden, die eine eindeutige Zuordnung und Platzierung eines bestimmten Images im Kontext ermöglichen. Die Dateinamen müssen sozusagen den Griff ins Bücherregal abbilden:

"Eine solche Strukturierung muss nicht nur die stimmige Seitenreihenfolge gewährleisten, sondern auch Kriterien zur Verfügung stellen, die eine Unterscheidung von Zeitschriften, Jahrgängen, Heften und Aufsätzen ermöglicht" (vgl. Ritzi 2000, S. 67).

Darüber hinaus müssen diese Strukturdaten nicht auf den Dateinamen beschränkt sein, sondern können im Rahmen der Erschließung je nach Erschließungskonzept (vgl. Kap.3.3) vor allem bei einer Volltextdigitalisierung Bedeutung erlangen.

### 3.2.4 Scannen der Presseauschnitte bei der Digitalisierung des Personenarchivs

Die Deutsche Forschungsgemeinschaft empfiehlt die im Rahmen eines Digitalisierungsprojektes anfallenden Teilaufgaben so weit wie möglich an externe Dienstleistungsfirmen zu vergeben. Daher wurden im HWWA bereits im Rahmen der Prüfung der Realisierbarkeit des Projekts diverse sondierende Firmenkontakte durchgeführt.

Im Gegensatz zur Digitalisierung im Haus (mittels Miete oder Kauf entsprechender Scanner und Bereitstellung eigenen Personals) hat die Vergabe des technischen Scannvorgangs an externe Dienstleister einen großen Vorteil:

„Ein Outsourcing der Scann-Dienstleistung erlaubt (...), jeweils den Dienstleister mit der passenden Erfahrung und dem passenden Equipment auszuwählen“ (Dörr 1999:592)

---

<sup>19</sup> CD-ROM = Compact Disc Read-Only Memory, bis zu 800 MB Speichervolumen

<sup>20</sup> DVD = Digital Versatile Disc, ursprünglich "Digital Video Disc"; bis zu 17 GB Speichervolumen bei doppelseitiger Nutzungsmöglichkeit.

Ein Outsourcing des Scannens setzt allerdings klar definierbare Aufgaben, Verfahren und Mengengerüste voraus, auf deren Basis Unternehmen Geräte- und Personaleinsatz und damit Preise kalkulieren können.

Daher wurde im HWWA ein Pflichtenheft erarbeitet, das die Menge und Besonderheiten des Materials benennt und Anforderungen an die zu liefernden Dateien festschreibt. Ein Ausschreibungsverfahren wurde durchgeführt. Die Auftragsvergabe erfolgte an einen Dienstleister, der den Scannvorgang übernimmt und Bilddateien im TIFF-Format auf dem Trägermedium DVD liefert. Die Images liegen als 300 dpi-Graustufen-Scann in Dateien vor, die nach einer vom HWWA vergebenen Nomenklatur benannt werden.

Da bei den gesammelten Presseausschnitten (im Gegensatz zu Büchern mit Seitenzahlen) keine eindeutige Identifikation einzelner Dokumente möglich war (die Artikel liegen zwar chronologisch geordnet, jedoch nicht durchnummeriert in den Sammelmappen) musste zunächst ein Verfahren entwickelt werden, das die Dokumente eindeutig identifiziert. Es sollte sich hierbei um eine Benennung handeln, die in den Datei-Namen aufgenommen werden soll und die zugleich einer späteren elektronischen Auswahl dienen kann.

Dazu wurde jedes Dokument mit einer Nummern-Kombination versehen, die aus drei Informationen-tragenden Teilen besteht: Neben einer Personen-ID, unter der die Stammdaten zu den Personen in der Datenbank geführt werden, besteht dieser Dateiname aus einer Dokumentenzählung und einem Sperrvermerk, der den urheberrechtlichen Status des Dokuments beschreibt: Das erste Dokument der Personenmappe mit der Personen-ID 01745 würde folgenden Dateinamen aufweisen: 01745-0001-000, wobei die letzten drei Zeichen den urheberrechtlichen Status des Dokuments beschreiben: hier eine urheberrechtlich unbedenkliche Agenturmeldung.

Diese „äußere“ Struktur der Presseausschnittssammlung lässt sich also durch den Dateinamen ausdrücken. Das Erschließungsinstrument NEWBASE<sup>21</sup>, das im HWWA auch für die aktuelle Presse genutzt wird, entnimmt den Dateinamen die drei enthaltenen Informationen (Personen-ID, Dokumentenzählung, urheberrechtlichen Status) automatisch und speichert diese als Metadaten.

---

<sup>21</sup> NEWBASE = Software für das Medien-Informations-Management; „NEWBASE deckt den kompletten Produktionsprozess für klassische Pressespiegel (Presseclipping) und digitale Medienspiegel ab und ist die Grundlage vieler erfolgreicher Geschäftsbeziehungen zwischen Informationslieferanten und –nachfragern“ ([www.newbase.de](http://www.newbase.de); 07.02.2005).

Die automatische Erfassung der inneren Struktur der Dokumente mit Hilfe einer „Szenenanalyse“ (die Elemente des Dokuments wie Überschrift, Autorenangabe, Bild, Textblock werden automatisch erkannt) ist für die aktuellen Presseartikel mit NEWBASE größtenteils möglich. Die Binnenstruktur der digitalisierten historischen Presseauschnitte kann jedoch wegen der höchst unterschiedlichen Struktur (auch im Vergleich zum Buch oder Zeitschriftenartikel) für diesen Altbestand nicht automatisch generiert werden. Solche Merkmale können nur manuell über eine Erfassungsmaske aufgenommen werden.

Als weiterer notwendiger Schritt der Vorbereitung der Dokumente für das Scannen außerhalb Haus wurden Verfahren für den Versand der Dokumente an den Dienstleister, sowie für deren Wiedereingliederung in das Archiv und für die Kontrolle der gelieferten Dateien entwickelt.

### **3.3 Inhaltliche Erschließung der Digitalisate**

In Bezug auf die inhaltliche Erschließung der Digitalisate müssen grundsätzliche Entscheidungen getroffen werden: Soll die inhaltliche Erschließung flach oder tief sein? Wie sollen die Seiten wiedergegeben werden? Als Bilder in Form von Blätterlösung oder als recherchierbarer Text? Wenn man sich für einen recherchierbaren Volltext entschieden hat, stellt sich die Frage, wie der Text erfasst werden soll: Kann der Text mittels einer OCR automatisch in ein binäres Format überführt werden? Oder müsste er manuell abgeschrieben werden? Und schließlich noch: Soll das originale Erscheinungsbild erhalten bleiben? Oder soll es durch eine reine Textdarstellung ersetzt werden?

#### **3.3.1 Erschließungskonzepte**

Die Deutsche Forschungsgemeinschaft beschreibt die Bedeutung der inhaltlichen Erschließung für die elektronische Bereitstellung von digitalisierten Dokumenten folgendermaßen:

„Inhaltliche Erschließung bedeutet, Dokumente eindeutig zu identifizieren, über ihren Inhalt zu informieren und Hilfestellung zur Beurteilung der Relevanz eines Dokumentes zu bieten. Sie kann in die Option münden, Zugang zum Dokument selbst zu erhalten. Diese Aufgaben sind für elektronische Publikationen neu zu definieren und zu organisieren, weil sie sich in einigen Positionen von den Voraussetzungen, Zielen und Methoden unterscheiden, wie sie der Erschließung von Druckschriften zugrunde liegen“ ([http://www.dfg.de/aktuelles\\_presse/reden\\_stellungnahmen/download/epub.pdf](http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/download/epub.pdf), 31.01.05)

Man kann für digitalisierte Bestände verschiedene Erschließungskonzepte unterscheiden (erweitert nach Brantl 2003:14f.):

- a) Will man in erster Linie Images (Faksimile der Dokumente) bereitstellen:
1. Zugriff entsprechend der physischen Vorlage: Blättern in den Images.
  2. Images und digitalisiertes Inhaltsverzeichnis: Das Inhaltsverzeichnis ist mit den Bildern verlinkt. Beispiel: Projekt „Bayerische Landtagshandlungen 1429-1669“<sup>22</sup> der Bayerischen Staatsbibliothek.
  3. Wie 2., aber zusätzlich mit digitalem Register: Das Inhaltsverzeichnis und die Register sind mit den Images verlinkt. Ein Beispiel hierfür bietet das Digitalisierungsprojekt „Reichstagsberichte 1867-1895“ der Bayerischen Staatsbibliothek<sup>23</sup>:

„(...): ein kumuliertes Sachregister (das sogenannte Generalregister), das 1895 erschienen ist und den gesamten Zeitraum abdeckt, wurde nach genauen Strukturierungsvorgaben des Digitalisierungszentrums von einem Dienstleister im Volltext erfaßt. Ergänzend wurden aus allen Einzelbänden die dort enthaltenen Sprecherregister digitalisiert. Diese wurden zwar nicht in einem gemeinsamen Index kumuliert, aber im Präsentationssystem kann über alle Bände hinweg gesucht werden“ (Dörr 2000:159).

4. Images und kompletter Text. Oft handelt es sich um einen unkorrigierten Text, der z.B. durch eine automatisierte Erfassung entstanden ist. Eine Volltextsuche im Text führt als Trefferanzeige immer nur zum Image. Ein Beispiel ist das Projekt der Deutschen Bibliothek in Frankfurt a.M. „Exilpresse digital. Deutsche Exilzeitschriften 1933-1945“<sup>24</sup> oder das Projekt „CAMENA“<sup>25</sup>:

„Diese Zweigleisigkeit bestimmt nicht nur die Herstellung, sondern auch die Nutzung des Textcorpus. Der Leser darf die Volltextversion nicht als Edition im üblichen Sinne ansehen. Er soll sie zur Recherche, zum Überfliegen bzw. schnellen Lesen und zur Weiterverarbeitung nutzen. Zur intensiven Lektüre aber wird er besser auf die zugehörigen Seitenabbildungen zurückgreifen. (...) Dies empfiehlt sich nicht nur wegen der planmäßig vorgenommenen Standardisierung, sondern auch wegen der bei dem gegebenen Verhältnis von Textmenge und Personalkapazität gewiss nicht geringen Fehlerhäufigkeit“ (Schibel/Kredel 2001:191).

- b) Will man primär Text bereitstellen:
1. Nur Text. Ein Beispiel ist das Projekt „Jahresberichte für deutsche Geschichte“ der Berlin-Brandenburgischen Akademie der Wissenschaften<sup>26</sup>
  2. Text und Images als Ergänzung, d.h. der Text wurde unter Verzicht auf bestimmte Textpartien erstellt. Um einen Informationsverlust zu vermeiden, wird

---

<sup>22</sup> <http://mdz.bib-bvb.de/digbib/bayern/byl/>

<sup>23</sup> <http://mdz.bib-bvb.de/digbib/reichstag>

<sup>24</sup> <http://deposit.ddb.de/online/exil/exil.htm>

<sup>25</sup> <http://www.uni-mannheim.de/mateo/camena.html>

<sup>26</sup> <http://www.bbaw.de/vh/jdg/digi.html>

die Buchseite als Faksimile hinterlegt. Beispiel: Projekt „Decretum Gratiani (Kirchenrechtssammlung)“<sup>27</sup> der Bayerischen Staatsbibliothek.

3. Kompletter Text, möglichst fein strukturiert. Alle strukturiert erfassten Textteile sind später gezielt recherchierbar. Beispiel: Projekt „Digitalisierung, Erschließung und Bereitstellung der Regesta Imperii 1831-1999 und künftiger Bände/Abteilungen“<sup>28</sup> der Deutschen Kommission für die Bearbeitung der Regesta Imperii e.V.

Technisch eine Stufe weiter und auch weit kostenaufwändiger ist, durch die coordierte Erfassung der Informationen in XML, einen zeichen- und vorlagentreuen Volltext zu reproduzieren.

- c) Hybridverfahren: Texte und Images ergänzen einander:

Als Maximum der bisherigen Möglichkeiten ist die hybride Digitalisierung anzusehen, die alle technischen Möglichkeiten parallel realisiert und Metadaten, Volltexte und Originallayout sowie z.B. zusätzlich noch einen Mikrofilm als dauerhaftes Speichermedium umfasst.

Bei vielen deutschen Digitalisierungsprojekten sind die Bemühungen darauf gerichtet, die Nähe zum Buch zu wahren. Daher wird die Image-Digitalisierung bevorzugt, bei der der Zugriff auf die Seitenbilder über Metadaten erfolgt, die z.B. aus Lemmalisten, Inhaltsverzeichnissen oder auch aus Registern bestehen, welche über Hyperlinks mit den entsprechenden Buchseiten verknüpft sind.

„So wird ein radikaler Medienbruch vermieden, doch zugleich der eigentliche Vorteil der elektronischen Medien kaum genutzt“ (Burch et. al. 2001:7).

Reine Blätterlösungen gelten heute als Minimallösung, denn je näher die digitale Version dem Buchmedium steht, umso weniger bietet sie mehr als das Buch selbst. Ziel sollte es sein dem Benutzer zu jeder digitalisierten Seite den entsprechenden Kontext zu präsentieren sowie entsprechende Navigationsinstrumente über den gesamten Inhalt zu bieten (vgl. Enders 2001:225). Dies lässt sich z.B. durch die Verlinkung mit Inhaltsverzeichnis und Registern verwirklichen.

Will man mehr als das Blättern in digitalen Bildern von Buchseiten anbieten, ist ein gezielter Zugriff auf die einzelnen Zeichen eines Textes eine notwendige Voraussetzung. Dies lässt sich durch die Volltextdigitalisierung schaffen, die „zwar einen große-

---

<sup>27</sup> <http://mdz.bib-bvb.de/digbib/gratian>

<sup>28</sup> <http://www.regesta-imperii.org/>

ren Aufwand bedeutet, aber den optimalen Weg eröffnet, die Vorteile des elektronischen Mediums umfassend auszuschöpfen“ (Burch 2001:7).

Stellt man jedoch ausschließlich den Text als OCR-Ergebnis bereit, fallen zwar einige Probleme wie eingeschränktes Navigieren oder die Dateigröße als Performancehinderung weg, allerdings verliert man nun einen guten Teil der Information des Druckoriginals wie die formale Gestaltung: Es erweist sich „das Mittel der Imagedigitalisierung gerade in den Geisteswissenschaften als besonders sinnvoll, weil in der Bildform oft dauerhaft benötigte Informationen enthalten sind, die durch seine Volltextfassung teilweise nicht wiedergegeben werden können“ (Mittler 1997:84).

Auf die Darstellung der Originalseite sollte also möglichst nicht verzichtet werden:

„Gerade für das wissenschaftliche Arbeiten ist die Authentizität eines elektronischen Dokuments von großer Bedeutung. Da jedoch jede Konvertierung in einen elektronischen Text statt in ein elektronisches Bild des Textes mit Informationsverlust verbunden ist, stellt das Image als digitales Faksimile die geeignetste Form dar, um die geforderten Authentizität zu erzielen“ (Enders 2001:226).

Zusammenfassend kann festgestellt werden, dass die Wahl des Konzepts in unmittelbarer Beziehung mit Erfassung und Bereitstellung steht, und damit zu den entstehenden Kosten: Das optimalste Erschließungskonzept, ein Hybridverfahren bei dem die Images der Seiten als Faksimile den vollständig erschlossen und mit XML strukturierten Text ergänzen, ist zugleich auch das kostenaufwändigste Verfahren.

„Die Hauptkosten für Digitalisierungsprojekte fallen durch die Erschließungstiefe und den dadurch bedingten Aufwand für die Erfassung und auch die Umsetzungen für die Bereitstellung an“ (Brantl 2003:15).

Die Wahl des Erschließungskonzepts stellt eine komplexe Entscheidung mit weitreichenden Folgen dar:

„Die Wahl der sinnvollsten Methode der Digitalisierung beschränkt sich also keineswegs auf ein einfaches Abwägen von Aufwand und Nutzen, vielmehr stehen bei einer gewissen Bandbreite von technischen Möglichkeiten und inhaltlichen Konzeptionen unterschiedliche Aufwände auf der einen Ebene und Einfach- bis Mehrfachnutzen auf der andere Ebene, beide zumindest teilweise miteinander verbunden und voneinander abhängig oder in Konkurrenz stehend“ (Leskien 2000:146).

### **3.3.2 Für und Wider Volltextdigitalisierung**

In Fällen, wo eine gut leserliche Antiqua-Vorlage zur Digitalisierung kommt, kann man durch eine elektronische Texterkennung (OCR) erheblichen Nutzen stiften, da die Dokumente nun auch im Volltext recherchiert werden können.

Aufgrund der weiten Verbreitung von Frakturschriften ist die Anwendung dieser Methode in Deutschland jedoch deutlich eingeschränkt:

„Es sei trotz Fortschrittsglauben die Prognose gewagt, dass eine hinreichende Genauigkeit und Verlässlichkeit derartiger Software-Tools wohl noch Jahre auf sich warten lassen wird“ (Leskien 2000:147).

Ein Beispiel bietet das Projekt zur „Digitalisierung der Reichstagsberichte 1867-1895“ an der Bayerischen Staatsbibliothek:

„Eine Volltextdigitalisierung war ausgeschlossen: die Reichstagsberichte sind in Fraktur gedruckt – auch trainierbare OCR-Programme führen bei dieser Schrifttype noch zu hohen Fehlerraten und eine manuelle Erfassung hätte angesichts des Umfangs ebenfalls außerhalb jeglicher Finanzierungsgrenze gelegen“ (Dörr 2000:158).

Ähnliche Probleme entstehen bei mathematischer, chemischer und physikalischer Formelerkennung mittels einer OCR, da in absehbarer Zeit nicht mit einer fehlerfreien Formelerkennung zu rechnen ist (vgl. Enders 2001:226f.), sowie bei hebräischen Texten.

Bisher sind die Ergebnisse der Digitalisierung mit automatischer Texterkennung bei solcher Art Vorlagen nur durch menschliche Ergänzungsarbeit nutzbar.

Der Einsatz von OCR kann - je nach Vorlage – neben den problematischen Fraktur-Schriften auch bei Antiqua-Schriften zu einer recht hohen Fehlerzahl führen, so dass die so erzeugten Volltexte für wissenschaftliches Arbeiten zunächst ungeeignet erscheinen. Ein denkbare Lösung hierbei sind z.B. doppelte bzw. mehrfache OCR-Läufe und der Abgleich ihrer Ergebnisse:

„Die manuelle Doppelerfassung mit automatischem Fehlerabgleich ist momentan die sicherste Methode für die Erfassung von Frakturschrift“ (Brantl 2003:15).

Auch bei nicht geeigneten Vorlagen kann ein OCR-Lauf hilfreich sein, weil er neben dem Ballast durch fehlerhafte Stichworte zumindest einige weitere Einstiegspunkte in den Dokumentinhalt bietet. Auch das Göttinger Digitalisierungs-Zentrum verfährt so:

„Da die Herstellung von Volltext nach unserer Ansicht sehr aufwändig und teuer ist, wird am GDZ darauf verzichtet. Der Benutzer soll lediglich die Seite als Faksimile (Image) präsentiert bekommen. Der erstellte Volltext dient im GDZ ausschließlich zum Retrieval“ (Enders 2001:227).

Insgesamt sollte bei einer Entscheidung über die Durchführung einer Volltextdigitalisierung bedacht werden, dass die Relevanzrate (Precision) bei der Recherche in Volltexten nicht so hoch ist. Allerdings erreicht man eine höhere Vollständigkeitsrate (Recall).

Bei einer tiefen intellektuellen Erschließung erreicht man eine höhere Genauigkeit bei der Recherche (hohe Precision), bei erheblich geringerem Ballast als bei einer Volltextrecherche.

Eine manuelle Erfassung von Volltext erscheint nur im Rahmen kleinerer Projekte sinnvoll, da sie sehr aufwändig und teuer ist. Die Arbeitsgruppe Technik empfiehlt, zumindest einzelne Strukturelemente eines Buches manuell zu erfassen und zu indizieren, wobei als Mindeststandard das Inhaltsverzeichnis gilt und zusätzlich das Register sehr empfohlen wird (vgl. Mittler 1998:57f.).

### **3.3.3 Erschließen der digitalisierten Presseauschnitte des Personenarchivs**

Die Erschließung der digitalisierten Presseartikel der Archive des HWWA gehört sicherlich zu den größten Herausforderungen dieses Digitalisierungsprojektes:

Die große Menge an einzeln zu erschließenden Dokumenten (ca. 120.000 im Teilprojekt Personenarchiv; ca. 3 Mio. im Gesamtprojekt) sowie die Art und Eigenschaften des Materials erfordern komplexe Problemlösungsstrategien und einen hohen Personaleinsatz: Die uneinheitliche Struktur und physische Größe der Dokumente sowie das uneinheitliche und teilweise aufgelöste Druckbild der Artikel erlauben keine automatisierte Erfassung, die bei einem solchen großen Bestand eine zügige Bearbeitung erleichtern würde.

Eine ursprünglich geplante Volltextrecherche lässt sich derzeit nicht realisieren, da die besonderen Merkmale der Vorlagen dies nicht zulassen: Die Zeitungsausschnitte sind neben Antiqua überwiegend in der Schrifttype Fraktur gesetzt. Innerhalb eines Artikels kommen teilweise verschiedene Frakturtypen zum Einsatz. Dazu kommt die große Vielfalt an Vorlagen (ca. 700 Quellen), die z.T. innerhalb des bearbeiteten Zeitraums ihre Schrifttypen mehrfach wechseln. Der Zustand der Vorlagen erschwert zudem eine erfolgreiche OCR: Sie sind wellig oder weisen Kontrastchwäche aufgrund der Bräunung des Papiers auf. Teilweise sind Zeichen als Folge steifen Papiers oder von Tintenfraß nur unvollständig erhalten. Bei Versuchen mit einigen OCR-Läufen wurden Fehlerquoten von 50% erreicht, die nicht hinnehmbar sind. Auch als erweiterter Sucheinstieg sind solcher Art erzeugte Stichwörter nicht nutzbar, da zuviel Ballast anfällt.

Allerdings ist im Hinblick auf zukünftige technische Entwicklungen, die eine Texterkennung auch von Frakturschriften ermöglichen könnten, in der Datenbank ein Modul integriert worden, über das zu einem späteren Zeitpunkt auf die Volltexterkennung mittels OCR zurückgegriffen werden kann.



Es muss jedoch auch zum jetzigen Zeitpunkt nicht vollständig auf eine OCR verzichtet werden: Denkbar ist, zunächst Dokumente für eine Volltexterkennung auszuwählen, bei der die Quelle in Antiqua gesetzt ist (nach der Erschließung können quer durch den Bestand alle Dokumente einer Quelle aufgerufen und mittels einer OCR behandelt werden). So könnten nach und nach die erfolgversprechendsten Vorlagen ausgewählt und das Programm auf die Anforderungen der speziellen Quelle hin trainiert werden. In Zukunft können bei verbesserten technischen Gegebenheiten auch die in Fraktur-Typen gesetzten Quellen so behandelt werden, so dass in Zukunft schrittweise eine Volltextrecherche im gesamten Bestand verwirklicht werden kann.

Aber auch ohne eine sofortige Volltextrecherchemöglichkeit wird durch die Digitalisierung und Erschließung ein erheblicher informationeller Mehrwert geschaffen: Über das Schubladenprinzip der Zuordnung der Dokumente zu einzelnen Personen und das chronologische Blättern in den Mappen hinaus, können durch die Erfassung der auf den Dokumenten enthaltenen Informationen (Personen-Name, Quelle und Ort, Erscheinungsdatum, Autor und Titel) verschiedenste Anfragen generiert werden:

- nach *Person*: einzige bereits im physischen Archiv verwirklichte Recherchemöglichkeit;
- nach *Autor*: alle Artikel eines bestimmten Autors<sup>29</sup>;
- nach *Erscheinungsdatum*: Alle Artikel zu einem bestimmten Ereignis, Tag oder in einem bestimmten Zeitraum;
- nach *Quelle*: Beobachtung einer bestimmten Zeitung / Zeitschrift. Automatisch mit dem Feld „Quelle“ verknüpft ist das Feld „Sprache“, bei dem alle Artikel, die in einer bestimmten Sprache verfasst sind, ausgewählt werden können.
- nach *Titel*: Über dieses Feld ist eine sachlich-inhaltliche Suche möglich, die weit über die vorherige Suchmöglichkeit über einzig den Personen-Namen hinausgeht.
- Kombinationen der Felder.

Die Merkmale „Personen-Name“, „Dokumentenzählung“ und „urheberrechtlicher Status“ werden automatisch aus dem Dateinamen der Images generiert (vgl. Kap.3.2.4.). Die Merkmale „Autor“ und „Quelle“ werden Vorschlagslisten entnommen, die bei der

---

<sup>29</sup> Ein weiterer informationeller Gewinn ließe sich m.E. mit einer Feld-übergreifenden Suche in den Feldern „Person“ und „Autor“ ermöglichen, da Personen auch in anderen Personenmappen als Autor auftreten können, was einem Nutzer allerdings vielleicht nicht sofort ins Auge springt. So kann vollständig Material „von“ wie „über“ eine Person gefunden werden. Ein Beispiel wäre die Person „Theodor Heuß“, über die sowohl eine Personenmappe angelegt ist, die zudem aber häufig in anderen Mappen als Autor auftritt.

Vorbereitung der Dokumente für das Scannen erstellt wurden. Die Titel – und im Falle fremdsprachiger Artikel deren Übersetzung – werden manuell erfasst. Darüber hinaus wird vermerkt, ob Fotos, Tabellen oder sonstige graphische Elemente enthalten sind und auch die Dokumentenart (Artikel, Aufsatz, Nekrolog, Meldung, Flugblatt, Denkschrift etc.) wird aufgenommen.

Die Erschließung bzw. die Vergabe dieser Metadaten wird mit dem Programm NEWBASE durchgeführt, das bereits in der Dokumentation der aktuellen Presse des HWWA zum Einsatz kommt. Damit verfährt das HWWA ähnlich wie das Projekt der Retrodigitalisierung der Akademieschriften der Königlich-Preußischen Akademie der Wissenschaften: Das historische Material wird mit Mitteln der Metadatenerzeugung der aktuellen Pressedokumentation behandelt:

„Grundidee hinsichtlich der Erzeugung von Metadaten ist dabei in diesem Projekt, den bisher gedachten Ansatz der Erzeugung von Metainformationen, der entweder auf die aktuelle Produktion von wissenschaftlichen elektronischen Dokumenten oder auf die nachträgliche ‚Veredlung‘ des Rohstoffs Information abhebt, konsequent auf die retrospektive Digitalisierung von Bibliotheksbeständen abzubilden. D.h. zu einer in digitalisierter Form vorliegenden Abhandlung werden mit dem zur Verfügung stehenden Instrumentarien (XML-Editor) Metainformationen in der Art gebildet, als handle es sich um eine „aktuelle“ elektronische Produktion“ (Wawra 2001:211).

Jedoch muss das Erschließungsinstrument NEWBASE an die besonderen Erschließungsanforderungen des Altmaterials angepasst werden. In vertrauter Arbeitsumgebung des Programms können in Zukunft – auch über die Dauer des Projekts hinaus – speziell für die Altbestände entwickelte Erfassungsmasken für die Erschließung benutzt werden.

„Langfristig entsteht auf diese Weise, wenn die noch bestehende Lücke in der Digitalisierung der Altbestände von den 1930er Jahren an bis zur gegenwärtigen Pressedokumentation geschlossen sein wird, ein einziges, homogenes Erschließungs- und zugleich Rechercheinstrumentarium, in das sämtliche je im HWWA entwickelten und genutzten Erschließungskonzepte integriert und insoweit verfügbar sind“ (aus dem Zwischenbericht, Huck; Striefler 2004:o.A.)

Da dieser Erschließungsprozess sehr zeitaufwändig ist und auch über die Projektdauer hinaus vom HWWA fortgeführt werden muss, ist es notwendig eine „Zwischenlösung“ zu entwickeln, um die digitalisierten Images auch kurzfristig präsentieren zu können:

Diese besteht zunächst in einer Blätterlösung, welche die Struktur der bisherigen physischen Archivierung abbildet: Anknüpfend an die in der Datenbank IFIS<sup>30</sup> geführten Personen-Stammdaten werden die Images konvolutartig den entsprechenden Personen

---

<sup>30</sup> IFIS = für das HWWA entwickeltes Internet-basiertes Fachinformationssystem

zugeordnet, chronologisch sortiert und, falls urheberrechtlich möglich, – entsprechend der Mappen im physischen Archiv – angeboten.

Im Gegensatz zum endgültigen Erschließungskonzept, das eine Erfassung auf Artikel-Ebene vorsieht, können hier durch eine Erschließung auf „Mappen-Ebene“ Metadaten den dossierartigen Zusammenstellungen von Artikeln zugeordnet werden. Der große Vorteil ist, dass es sich hierbei um „lediglich“ 3.200 zu den jeweiligen Personen gehörende Datensätze handelt (ca. 8.000 im Falle des Firmenarchivs), denen von einem Artikel bis zu Hunderten von Artikeln zugeordnet sein können.

Bereits bei der Vorbereitung des Materials zum Scannen wurden zum Teil folgende sachlich-inhaltlichen und bibliographisch-formalen Merkmale erhoben und in die Datenbank IFIS überführt, in der die Stammdaten zu den Personen (und im Falle der Digitalisierung des Firmenarchivs auch Stammdaten zu Firmen) gepflegt werden. Bei den Personen handelt es sich um folgende Daten: Personen-Name (Ansetzung entsprechend PND<sup>31</sup>), Verweisformen, Lebensdaten, Wirkungsbereich, Herkunfts- und Wirkungsland, z.T. Kurzbiographien, Auszeichnungen und Preise, weiterführende URLs, Materialbestand von/bis, Person als Autor (ja/nein), Abbildungen.

Auf der Ebene der Dossiers bzw. Mappen könnten durch diese tiefergehende Erschließung zudem weitere Produkte entwickelt werden:

Durch eine Abfrage nach Beruf und Wirkungsland könnten z.B. alle britischen Wirtschaftswissenschaftler oder alle Staatsoberhäupter, alle Kolonialbeamte oder alle europäischen Politiker zusammengestellt werden. Denkbar sind auch Zusammenstellungen von z.B. Nobelpreisträgern oder religiösen Würdenträgern. Kombinierbar sind sachliche, geographische und auch zeitliche Aspekte. Das Ergebnis einer Recherche wären hierbei immer Gruppen von Dokumenten.

Auf dieser Ebene könnten auch bestandsübergreifende semantische Verknüpfungen hergestellt werden. So könnten z.B. Firmenchefs aus dem Bestand des Personenarchivs mit den jeweiligen Firmen im Firmenarchiv und dem entsprechenden Produkt im Sacharchiv verknüpft werden. Allerdings sind diese Verknüpfungen nahezu ausschließlich intellektuell zu erstellen, da keine Suchfrage diese automatisch generieren könnte.

Es sind neben text-basierten Abfragemöglichkeiten auch visuelle Rechercheangebote denkbar: Etwa eine visuelle Aufbereitung in Form einer geographischen Suchfunktion

---

<sup>31</sup> PND = Personennamendatei

wie z.B. auch bei dem Projekt „Digitalisierung des historischen Kartenbestands der SuUB Bremen“<sup>32</sup> in Form einer anklickbaren Karte. Insbesondere für die Dokumente des Firmenarchivs könnte eine anklickbare Karte der Firmensitze eine geographische Recherche erleichtern: Bei einer text-basierten Suche müssten aufgrund historischer Grenz-Verschiebungen unterschiedliche nationale Bezeichnungen in die Recherche einbezogen werden; auf der Karte handelt es sich jedoch um einen fest definierten Punkt. Auch auf der Ebene der Artikel könnte durch die Erschließung eines geographischen Ortes eine Bestands- und Mappen-übergreifende visuelle Suche ermöglicht werden.

Zusammenfassend kann festgestellt werden, dass aufgrund der Besonderheit der Bestände des HWWA – die dossierartige Zusammenstellung von Mappen – ein dualer Weg der Erschließung sinnvoll ist: Für die kurzfristige Bereitstellung sollte eine Recherche auf Ebene der Dossiers bzw. Mappen ermöglicht werden, wozu eine tiefergehende Erschließung der Stammdatensätze zu Personen (und auch Firmen) nötig wäre. Hieraus könnten weitere Produkte entwickelt werden. Jedoch sollte man das langfristige Ziel der Einzellerschließung der Artikel nicht aus den Augen verlieren, denn nur sie ermöglicht eine weitestreichende Generierungen von informationellem Mehrwert.

### **3.4 Bereitstellung und Präsentation der digitalisierten Dokumente**

Grundsätzliche Überlegungen zur Bereitstellung beinhalten zunächst die Frage, ob das Material Online oder Offline angeboten werden soll: Soll es im Internet bereit gestellt oder auf CD-ROM oder DVD veröffentlicht werden? Die Präsentation der Ergebnisse einer Digitalisierung auf CD-ROM und DVD ist aufgrund der vergleichsweise geringen Speicherkapazität hauptsächlich für text-basierte und nicht sehr umfangreiche image-basierte Digitalisierungsergebnisse zu nutzen. Im Folgenden werde ich ausschließlich die Bereitstellung über das Internet behandeln.

#### **3.4.1 Bereitstellung: technische Infrastruktur und inhaltliche Aspekte**

Um digitalisierte Seiten im entsprechenden Kontext zu präsentieren, bedarf es spezieller internet-basierter Dokument-Management-Systeme (DMS), die sowohl Inhalte als Images speichern und darstellen als auch die entsprechenden Metadaten und evtl. Strukturdaten aufnehmen können. Das Verwaltungssystem muss zudem Suchfunktionen und Zugriffsmöglichkeiten bieten.

„Im DMS des Göttinger Digitalisierungszentrum gehört zu jedem Image außerdem noch ein als elektronischer Volltext erstellter Metadatensatz, der sowohl Meta- als

---

<sup>32</sup> <http://gauss.suub.uni-bremen.de/suub/hist/index.jsp>

auch Strukturdaten enthält, die sowohl für das Retrieval als auch für die Navigation (Blätter, gezieltes Aufschlagen, elektronisches Inhaltsverzeichnis) genutzt wird“ (Enders 2001:226).

Die Anforderungen an ein solches Dokument-Management-System für die Verwaltung der Digitalisate sind (vgl. Dörr 1999:594):

- Nachbildung von Dokumentstrukturen für eine komfortable Navigation im Dokument;
- strukturierte Suche;
- Volltextsuche;
- flexible Anpassung von Darstellung und Suchmasken an jeweilige Sammlungen;
- Zusammenstellung von Sammlungen;
- Suche über Sammlungen hinweg sowie
- Import- und Exportschnittstellen für strukturierte Daten (XML als Format für Migration und Langzeitarchivierung).

Der Zugriff auf die Digitalisate kann sowohl aus dem DMS direkt geschehen oder auch über einen OPAC<sup>33</sup> erfolgen. Dazu müssen die digitalen Dokumente mit den im OPAC aufgeführten Metadaten verknüpft werden.

Zusätzlich können die Dokumente auch in überregionalen Katalogen und Suchmaschinen recherchierbar sein, vorausgesetzt sie sind anhand zugeordneter URLs<sup>34</sup> eindeutig identifizierbar.

Auf die meisten Retrodigitalisierungsprojekte kann über die Homepage der ausführenden Institution zugegriffen werden. Darüber hinaus sollte eine Einbindung der Ergebnisse in eine einheitliche Infrastruktur für eine „Verteilte digitale Forschungsbibliothek“ Ziel aller Projekte des Programms „Retrospektive Digitalisierung von Bibliotheksbeständen“ sein (vgl. Mittler 1998:66f.).

Sinnvoll ist daher, einen Zugriff auch über andere Anbieter zu ermöglichen. Bei dem Projekt „Retrospektive Digitalisierung des historischen Kartenbestandes der SuUB Bremen“ werden die Karten neben der Bereitstellung auf einer eigenen Homepage auch in der Online-Datenbank PICA und in der IKAR-Altkartendatenbank angeboten, die über die Homepage des GBV abgerufen werden kann (vgl. <http://gauss.suub.uni-bremen.de/suub/hist/index.jsp>; 07.02.2005).

Diese Einbindung ist mitentscheidend für den Erfolg eines Digitalisierungsprojektes:

---

<sup>33</sup> OPAC = Online Public Access Catalogue = Bibliothekskatalog im Internet

<sup>34</sup> URL = Uniform Resource Locator = Web-Adresse im Internet

„[...] umso wichtiger ist schrittweises zielgerichtetes Handeln auf einem langen Weg der Profilbildung einer digitalen Sammlung. Nicht zu unterschätzen ist in diesem Zusammenhang die Einbindung in das gesamte fachspezifische Informationsangebot, das als Management der Medienbrüche bezeichnet werden kann“ (Leskien 2000, S.142).

Insgesamt bedeutet dies, dass die Bereitstellung im Internet nur erfolgreich sein kann, wenn die Dokumente entsprechend vermarktet werden und leicht aufzufinden sind. Eine Veröffentlichung auf einer eigenen Homepage reicht hierfür meist nicht aus. Es empfiehlt sich daher, zusätzlich Zugriff über z.B. entsprechende Fachportale zu verwirklichen.

### **3.4.2 Präsentation des digitalisierten Personenarchivs des HWWA**

Die Präsentation des digitalisierten Personenarchivs des HWWA soll zunächst über zwei Wege erfolgen: Erstens über eine eigens erstellte Homepage sowie eventuell über den OPAC<sup>35</sup> des HWWA.

Die Präsentation wird dabei aus folgenden Datenquellen gespeist (vgl. Anhang 2):

Die Datenbank IFIS liefert die Stammdaten zu Personen (und auch Unternehmen im Fall des Firmenarchivs).

Die Images (Faksimile) liegen auf einem Server in drei verschiedenen Zooms (30%, 60%, 100% des Originals) als JPG vor. Hier werden in Zukunft auch die im Original in Übergröße (>DIN A4) vorliegenden Presseartikel im PDF-Format gespeichert und abrufbar sein.

Die Metadaten werden in der über das Internet zugänglichen, der HWWA-eigenen EDV-Infrastruktur angepassten Datenbank NEWBASE erzeugt und bereitgestellt. Langfristig sollen die Daten in den OPAC eingespeist werden. Dazu müssen Recherchefunktionen modifiziert und Bestellkomponenten konzipiert werden.

Für eine zeitnahe Präsentation wurde bereits ein erster Entwurf einer Webseite entwickelt (vgl. Anhang 3). Die Dokumente des Personenarchivs sind dort in einem alphabetischen Register oder mit Hilfe eines Suchfeldes nach Personennamen abfragbar. Als Ergebnis werden Verweise auf die Dokumente in Form von Bündeln in einer chronologisch sortierten Liste angezeigt. In Zukunft soll ein Blättern direkt von Artikel zu Artikel ermöglicht werden.

---

<sup>35</sup> Die Zukunft dieses Bereitstellungsinstrumentes ist noch nicht endgültig geklärt.

Die Dokumente liegen dort zunächst ausschließlich als Images vor. Bisher sind Artikel in Übergröße in mehrere Teilbilder von je DIN A4 zerschnitten. Dies, um einen schnellen Zugriff und – in Bezug auf die Skalierung - eine gute Lesbarkeit zu erreichen. Artikel in Übergrößen haben größere Ladezeiten und sind aufgrund der Skalierung teilweise nicht zu lesen. Der Nachteil ist jedoch, dass durch das Zerschneiden die Lesereihenfolge der in Spalten gesetzten Artikel beim Blättern nicht mehr eingehalten wird, so dass der Nutzer mehrfach zwischen den Bilddateien (Seiten) springen muss. Daher sollten im Zuge der Erschließung ergänzend mit Hilfe des Programms NEWBASE PDFs erzeugt werden (vgl. Kap. 3.3.1.). Diese entsprechen zwar nicht dem Layout des Originals, weisen aber die Vorzüge von A4-Seiten und zudem eine richtige Lesereihenfolge auf. Wie ich in Kap. 3.3.1 gezeigt habe, sollte aber auf die Präsentation der Faksimile auch bei offensichtlichen Nachteilen nicht verzichtet werden, da im Original-Layout oft dauerhaft benötigte Informationen enthalten sind.

Über die Präsentation in OPAC und Homepage hinaus sollte eine Zusammenarbeit mit weiteren Fachportalen ins Auge gefasst werden. Denkbar wären für den historischen Presseartikel-Bestand z.B. das Fachportal für Geschichtswissenschaften „CLIO online“<sup>36</sup> der Humboldt-Universität Berlin, das Geschichtsportale „Chronicon“ der Bayerischen Staatsbibliothek<sup>37</sup>, das interdisziplinäre Fachportal „VASCODA“<sup>38</sup>, eventuell die Virtuelle Fachbibliothek Politikwissenschaften „ViFaPol“<sup>39</sup> der SUB Hamburg, sowie das „Institut für Zeitungsforschung“<sup>40</sup> der Stadt Dortmund. Dazu müssen Fragen der Datenübergabe und -einbindung geklärt werden. Sinnvoll wäre die Übergabe der Metadaten im XML-Format mit Link zu den Daten des HWWA-Servers, auf den auf http-Anfrage zugegriffen werden kann.

### **3.5 Exkurs: Urheberrechtsproblematik**

#### **3.5.1 Retrodigitalisierungsprojekte und das Urheberrecht**

Da die Urheberrechtsproblematik an verschiedenen Stellen in den Prozess der retrospektiven Digitalisierung eingreifen und in Einzelfällen eine besondere Bedeutung erlangen kann, werde ich diesen Aspekt im Folgenden untersuchen.

---

<sup>36</sup> <http://www.clio-online.de/>

<sup>37</sup> <http://www.chronicon.de>

<sup>38</sup> <http://www.vascoda.de>

<sup>39</sup> <http://www.vifapol.de/>

<sup>40</sup> <http://www.zeitungsforschung.de/>

Bereits bei der Bestandsauswahl können urheberrechtliche Fragen auf die Wahl der Bestände Einfluss nehmen. So können urheberrechtlich bedenkliche Bestände ganz von der retrospektiven Digitalisierung ausgenommen werden.

Wie sich an den für die Digitalisierung im Rahmen des DFG-Förderprogramms ausgewählten Beständen zeigt, wird bei den meisten Projekten älteres, bereits urheberrechtlich freies Material (Ostraka, Papyri, mittelalterliche Handschriften etc.) für die retrospektive Digitalisierung ausgewählt.

Bei jüngerem Material aus dem 20. Jh. kann die Frage des Urheberrechts eine große Rolle für den Arbeitsprozess der Digitalisierung spielen. Dabei ist zu unterscheiden, ob ganze Bestände oder nur Teile von Beständen urheberrechtlich problematisch sind.

Urheberrechtlich problematisch sind Werke, deren Autor noch nicht seit 70 Jahren verstorben ist, denn hiernach erlischt nach deutschem Recht das persönliche Urheberrecht<sup>41</sup>.

Hat man sich für die Digitalisierung von Dokumenten entschieden, die unter das Urheberrecht fallen, so muss das ausgewählte Material auf seine Tauglichkeit zur Bereitstellung geprüft und eine Entscheidung für den Arbeitsprozess gefällt werden.

Es gibt hierbei verschiedene Vorgehensweisen:

1. aus dem Bestand das problematische Material entnehmen und von der Digitalisierung ausnehmen;
2. eine Einzelfallprüfung vornehmen und das Material bis zu seiner urheberrechtlich unbedenklichen Verwertungszeit sperren;
3. eine Einzelfallprüfung vornehmen und die Rechte von Verlegern und Urhebern einholen, um somit den gesamten Bestand bereits während der Projektphase frei verfügbar bereitstellen zu können.

Die DFG dachte im Rahmen ihrer Förderprogramme von vornherein daran, Rechteinhaber wie Autoren und Verleger an dem Vorhaben zu beteiligen, um wichtige, verwaltungsrechtlich geschützte Dokumente nicht ausgrenzen und damit den Wert von Digitalisierungsprogrammen erheblich mindern zu müssen (vgl. Leskien 2000:132). Das kann im Einzelfall z.B. die Bereitstellung von Projektmitteln für eine rechtsanwaltliche Beratung oder ein Rechtsgutachten bedeuten.

---

<sup>41</sup> Die Dauer von 70 Jahren wurde im Zuge der europäischen Harmonisierung in das europäische Recht übernommen. Eine Ausnahme stellen allerdings Veröffentlichungen aus Spanien dar, wo (noch) eine Dauer von 80 Jahren für das persönliche Urheberrecht gilt.

Das Urheberrecht online: <http://bundesrecht.juris.de/bundesrecht/urhg/>; 15.01.2005



Auch beim Retrodigitalisierungsprojekt des HWWA wurde ein einschlägig qualifizierter Rechtsanwalt herangezogen, der für das Projekt ein Rechtsgutachten erstellte, das sich neben urheberrechtlichen Fragen auch mit Fragen des Persönlichkeitsrechts und weiteren eventuell berührten Rechtsbereichen auseinandersetzt.

### **3.5.2 Die Urheberrechtsproblematik bei der Digitalisierung des Personenarchivs**

Die Digitalisierung von Artikeln aus einem Pressearchiv und deren Bereitstellung im Internet sind – ohne Einigung mit dem Urheber – nicht erlaubt, da sie eine Vervielfältigungs-, Veröffentlichungs- und Verbreitungshandlung sowie eine öffentliche Wiedergabe darstellen und somit die Schutzrechte des Urhebers verletzen.

Die Ausnahmeregeln der §49 UrhG (Zeitungartikel und Rundfunkkommentare) und §53 UrhG (Vervielfältigung zum privaten und sonstigen Gebrauch) beschränken zwar die Rechte des Autors im Interesse der Allgemeinheit<sup>42</sup>, doch durch ein Urteil des Bundes-Gerichtshofs 1998 und ein Gutachten des Bundes-Ministeriums der Justiz wurde verneint, dass die Digitalisierung eine zulässige Vervielfältigung im Sinne der § 49 und § 53 UrhG ist (vgl. Beger 2000:o.A.).

Für die Digitalisierung des Personenarchivs des HWWA wurde Material von ca. 1880 bis 1930 aus internationalen Quellen<sup>43</sup> ausgewählt. Es zeigte sich, dass es sich bei einem Großteil der Artikel um urheberrechtlich unbedenkliche Agenturmeldungen handelt, doch insbesondere die ausführlichen und inhaltlich besonders relevanten Artikel oft namentlich gezeichnet sind. Ein Verzicht auf diese Artikel wäre nicht im Sinne des Projektes. Eine umfassende Klärung der Rechte mit den entsprechenden Verlagen bzw. ihren Rechtsnachfolgern wäre allein bei 700 deutschen und internationalen Quellen schon eine große Aufgabe geworden, doch da das Urheberrecht für diesen Zeitraum bei

---

<sup>42</sup> § 49, Abs. 1: Pressespiegel (auch Presseauschnittssammlungen) dürfen ohne Zustimmung einzelne Beiträge aus Zeitungen und Informationsblättern (20 %) enthalten, der Pressespiegel darf veröffentlicht werden, es besteht eine Vergütungspflicht an die VG Wort

§ 49, Abs. 2: Pressespiegel von Nachrichten dürfen ohne Zustimmung und Vergütung hergestellt und veröffentlicht werden.

§ 53, Abs. 2, Nr. 2 Archivkopie/ Pressearchiv: Zu Archivzwecken darf, soweit ein eigenes Werkstück benutzt wird, eine vollständige Kopie hergestellt werden.

§ 53, Abs. 2, Nr. 4a, sonstiger eigener Gebrauch: Einzelne Beiträge (20% aus einer Zeitung und Zeitschrift) dürfen vervielfältigt werden. Diese dürfen nicht der Öffentlichkeit zugänglich gemacht werden.

§ 53, Abs. 2, Nr. 4 b: Dieser ermöglicht Zeitungen, die seit mindestens zwei Jahren vergriffen sind, vollständig zu vervielfältigen, auch wenn kein eigenes Werkstück als Vorlage vorhanden ist.

Das Vervielfältigungsverfahren wird im Gesetz nicht bestimmt, so dass grundsätzlich alle Verfahren zulässig sind.

<sup>43</sup> Internationale Quellen sind ebenso wie deutsche Quellen zu behandeln: Nach internationalen Abkommen, wie der Revidierten Berner Übereinkunft, wird bei der Nutzung urheberrechtlich geschützter Werke das Prinzip der Inländerbehandlung angewandt (vgl. Beger 2000:o.A.).

den Autoren liegt<sup>44</sup> ist ein solcher Weg im Rahmen des Projektes generell nicht mehr leistbar: Es müssten zu mehreren Tausend Autoren und Fotografen Rechtsnachfolger ermittelt und Genehmigungen eingeholt werden.

Daher hat man sich im Projekt auf ein Vorgehen im Sinne einer „moving wall“ festgelegt. Ein Verfahren wie es ähnlich beispielweise auch das US-amerikanische Zeitschriften-Digitalisierungsprojekt „JSTOR“<sup>45</sup> nutzt:

„JSTOR bietet den Zugriff zu den ein bis fünf Jahre zurückliegenden Jahrgängen einer Zeitschrift. Der zeitliche Abstand zwischen den aktuellen Ausgaben einer Zeitschrift und der von JSTOR angebotenen älteren Jahrgängen wird dabei zwischen den Verlegern, die das Urheberrecht an den Zeitschriften besitzen, und JSTOR individuell vereinbart. Nach Ablauf eines Jahres wird von JSTOR ein weiterer Jahrgang eingearbeitet, wobei es sich bei der einen Zeitschrift um einen fünf Jahre zurückliegenden Jahrgang bei einer anderen Zeitschrift um einen nur zwei Jahre zurückliegenden Jahrgang handeln kann. Von JSTOR wird das Verfahren als „moving wall“ bezeichnet“ (Hilz 1999:221).

Dieses Verfahren bietet die Möglichkeit der Aktualisierung. Bei der Retrodigitalisierung der Pressearchive können durch die Nutzung dieses Prinzips urheberrechtsfrei gewordene Dokumente nach und nach zur Nutzung freigegeben werden:

Die auf den Artikel verzeichneten Autoren werden einer Einzelfallprüfung unterzogen. Da ihre Namen bereits im Zuge der Erschließung manuell aus den Artikeln entnommen werden, können bei diesem Arbeitsschritt ergänzend ihre Lebensdaten ermittelt werden. Dies geschieht mittels einer Recherche in Online-Bibliotheks-Katalogen (OPACs) und der Personennamen-Datei (PND), in Spezialdatenbanken (u.a. „Deutsches Historisches Museum“; „Koloniallexikon“; „Lexikon der deutschen Generäle“; „Rulers“) und mittels Internet-Suchmaschinen (z.B. Google). Anhand des Sterbedatums können Aussagen über ihren urheberrechtlichen Status gemacht werden: Sind die Autoren bereits seit mehr als 70 Jahren verstorben, können ihre Artikel bereitgestellt werden. Alle Artikel mit Autoren, die innerhalb der vergangenen 70 Jahre verstorben sind, bekommen einen Sperrvermerk, der ihr Todesjahr enthält, so dass diese Artikel, dem Prinzip einer „moving wall“ folgend, sukzessive bereitgestellt werden können. Artikel, bei den das Todesjahr des Autors nicht ermittelt werden konnte, müssen für eine Bereitstellung in absehbarer Zeit im Rahmen der „moving wall“, gänzlich gesperrt werden und erhalten einen gesonderten Sperrvermerk.

---

<sup>44</sup> „Das Recht zur Digitalisierung liegt bei vor 1995 erschienen Zeitschriften- und Zeitungsbeiträgen in der Regel beim Autor und nicht beim Verleger“ (Beger 2002:36)

<sup>45</sup> <http://www.jstor.org>

Wie ich gezeigt habe, stellt die Urheberrechtsproblematik eine besondere Herausforderung bei der retrospektiven Digitalisierung von historischen Zeitungs- und Zeitschriftenausschnitten dar. Die hohe Anzahl verschiedenster Quellen (700) in einem urheberrechtlich nicht unproblematische Zeitraum (1880-1930) in Kombination mit der Eigenschaft einer Zeitung bzw. Zeitschrift als Quelle, nämlich im Gegensatz zu Monographien eine Vielzahl von Autoren aufzuweisen, bedeutet einen hohen Aufwand für die Einzelfallprüfung. Der informationelle Gewinn, der dadurch entsteht, dass man das urheberrechtlich bedenkliche Material nicht aussondert, rechtfertigt jedoch den hohen Aufwand.

#### **4 Rückblick und Ausblick: Die besonderen Herausforderungen bei der Digitalisierung des historischen Personenarchivs des HWWA**

Wie ich in vorliegender Arbeit gezeigt habe, stellen die unterschiedlichen Medien und Gegenstände der Projekte zur „retrospektiven Digitalisierung von Bibliotheksbeständen“ die durchführenden Institutionen vor die unterschiedlichsten Herausforderungen in Bezug auf Bestandsauswahl, Scannen, Erfassen, Erschließen und Bereitstellen. Erschwerend kommt die enge Verzahnung dieser Gebiete im Arbeitsprozess hinzu:

„Die drei Bereiche Erfassung, Erschließung und Bereitstellung sind so eng miteinander vernetzt, dass nachträgliche Veränderungen in den Anforderungen an eine der Komponenten zu erheblichem Nachbearbeitungsaufwand in den anderen Modulen führen können“ (Brantl 2003:14).

Die von Projekt zu Projekt unterschiedliche Menge des zu digitalisierenden Materials und die Vielfalt an Quellen nimmt einerseits auf den Arbeitsprozess des Scannens wie andererseits – im Falle des Mediums Text – auch auf die nachfolgend evtl. gewünschte OCR Einfluss. Die (technisch) möglichen Ergebnisse der OCR wiederum beeinflussen die Wahl des Erschließungskonzeptes und somit auch der Gewinnung von Mehrwert. Die Wahl der Mittel bzw. Strategien hängt zudem sehr von den Entwicklungen in den Bereichen Hard- und Software ab, insgesamt also von den technischen Möglichkeiten und Rahmenbedingungen und ihren dynamischen Veränderungen.

Ein weiterer Problemkreis ist der Umgang mit dem Urheberrecht. Je nach Alter und Herkunft der Quellen kann der urheberrechtliche Schutz von Dokumenten Probleme bereiten, insbesondere, wenn die Anzahl an verschiedenen Rechteinhabern groß ist.

Diese Vielfalt an Eigenschaften, die in unterschiedlichen Kombinationen jeweils sehr spezielle Wirkung entfalten, führt zu sehr komplexen und speziellen Problemlösungsstrategien, wie ich im Verlauf dieser Arbeit beispielhaft für das Projekt der retrospektiven Digitalisierung der Pressearchive des HWWA gezeigt habe.

Die herausragende Besonderheit dieses Projekts ist zunächst in seiner Einzigartigkeit zu sehen. Bislang befasst sich kein weiteres Projekt mit der retrospektiven Digitalisierung von historischen Presseausschnittssammlungen. Zudem ist der Umfang des Projekts mit einer Zahl von etwa 3 Mio. Dokumenten bisher unerreicht.

Die größte Herausforderung des Gesamt-Projektes zur retrospektiven Digitalisierung der historischen Pressearchive des HWWA ist daher die im Vergleich zu anderen Projekten enorme Menge an Dokumenten. In Kombination mit dem gewünschten Erschließungskonzept - einer zunächst gewünschten Volltextdigitalisierung auf Artikelebene - verstärkten sich Probleme, die durch die Eigenschaften des Bestandes hervorgerufen werden:

- technische Probleme der automatischen Texterfassung:
  - Eine OCR ist aufgrund der Frakturschriften und des schlechten Druckbildes nicht möglich. Bei einem Projekt mit nur wenigen Tausend Dokumenten wäre eine manuelle Erfassung der Volltexte denkbar, bei 3 Mio. Dokumenten ist dies unmöglich.
  - Wäre die Vielfalt der Quellen eingeschränkt, könnte selbst bei großen Dokument-Mengen ein OCR-Programm trainiert werden. Bei ca. 700 verschiedenen Quellen, die ihre Schrifttypen im Verlauf der Jahre mehrfach wechseln und deren Druckbild von Tag zu Tag schwankt, müssen ebenso viele OCR-Profile trainiert werden.
- urheberrechtlicher Status: Bei den Presseartikeln handelt es sich – im Gegensatz zu den Gegenständen der meisten anderen Retrodigitalisierungsprojekte – um urheberrechtlich relevantes Material. Anders als bei historischen Drucken, Handschriften oder Bildern aus dem 19. Jh. und früher muss der urheberrechtliche Status der einzelnen Dokumente in eine Einzelfallprüfung erhoben werden. Bei anderen Digitalisierungsprojekten, deren Gegenstand aus dem 20. Jh. stammt, ist eine Klärung des urheberrechtlichen Status nicht so problematisch, da die Vielfalt an Quellen deutlich geringer ist (vgl. z.B. Projekt "Exilpresse digital. Deutsche Exilzeitschriften 1933-1945").
- zeitnahe Bereitstellung: Die außerordentlich große Menge an Dokumenten kann im Projektzeitraum nicht vollständig bearbeitet werden, so dass Übergangslösungen ergänzend zum endgültigen Erschließungskonzept erarbeitet werden müssen

Eine weitere besondere Herausforderung ist der zeitliche Wandel der Erschließungsinstrumente. Dies gilt allerdings weniger für die Digitalisierung des Personenarchivs

(bzw. des Firmenarchivs), das ausschließlich über den Personennamen (bzw. Firmennamen) erschlossen ist, als vielmehr für das Sach-, Länder und Warenarchiv, das jedoch mit ca. 2,7-2,9 Mio. Dokumenten das größte Teilprojekt darstellt:

Die Dokumentationssprachen (Klassifikation und Systematik) unterlagen in den vergangenen Jahren mehrfachen Veränderungen, so dass es notwendig wurde, eine Konkordanz zwischen den historischen und gegenwärtigen Erschließungsinstrumenten (Standard Thesaurus Wirtschaft) zu erarbeiten. Diese Aufgabe ist umso mehr von Bedeutung, als durch den mehrfachen Generationenwechsel der Mitarbeiter der Pressedokumentation des HWWA das Wissen über die alte Systematik verloren geht.

Zusammenfassend kann man sagen, dass durch die Digitalisierung des Personenarchivs bereits wichtige Lösungsstrategien für das Gesamtprojekt entwickelt werden konnten, die nach nur geringen Modifikationen auch für die noch durchzuführenden Teilprojekte Firmen- sowie Sach-, Länder- und Warenarchiv Anwendung finden können.

Über die Zeitdauer des Projekts hinaus muss das HWWA Anstrengungen unternehmen, die Erschließung der Artikel zu Ende zu führen, wobei der duale Weg der Erschließung auf Mappen-Ebene sowie auf Artikel-Ebene beschränkt werden sollte. Auch die Einbindung der digitalisierten Dokumente in weitere Fachinformationssysteme, Portale etc. ist eine noch zu lösende Aufgabe, die jedoch entscheidend für den Erfolg des Projektes sein kann. Zudem wäre es sinnvoll eine Fortschreibung des Projekts über den bisher ausgewählten Zeitraum (bis 1930) hinaus ins Auge zu fassen.

Durch weitere Kooperationen mit anderen Archiven könnten diese ermutigt werden, ihre eigenen historischen Bestände zu digitalisieren und so auf den Aufbau eines gemeinsamen digitalen historischen Pressearchivs hinzuarbeiten.

## 5 Literatur

- Beger, Gabriele (2000): Urheberrechtliche Würdigung von elektronischen Pressespiegeln und Archiven (unveröffentlichtes Arbeitspapier vom 09.11.2000)
- Burch, Thomas et al. (Hrsg.) (2001): Standards und Methoden der Volltextdigitalisierung. Beiträge des Internationalen Kolloquiums an der Universität Trier, 8./9. Oktober 2001. (Abhandlungen der Geistes- und sozialwissenschaftlichen Klasse Einzelveröffentlichung Nr. 9). Mainz: Akademie der Wissenschaften. Stuttgart: Franz Steiner Verlag
- Brantl, Markus: Konversion historischer Texte in digitale Medien. In: Mitteilungen aus dem Bundesarchiv, 11 (2003), Heft 1, S. 13-18
- Dörr, Marianne: Das Münchener Digitalisierungszentrum an der Bayerischen Staatsbibliothek. In: Bibliotheksforum Bayern, 28 (2000), Heft 2, S. 154-177
- Dörr, Marianne: Das Münchener Digitalisierungszentrum an der Bayerischen Staatsbibliothek. In: Bibliotheksdienst, 33 (1999), Heft 4, S. 592-600
- Effinger, Marja et al.: Per Mausclick ins Spätmittelalter: Digitalisierung und Erschließung spätmittelalterlicher Bilderhandschriften aus der Bibliotheca Palatina. In: B.I.T. online, 6 (2003), Heft 3, S. 235-248
- Enders, Markus: Volltextverarbeitung am Göttinger Digitalisierungszentrum. In: Burch, Thomas et al. (Hrsg.) (2001): Standards und Methoden der Volltextdigitalisierung. Beiträge des Internationalen Kolloquiums an der Universität Trier, 8./9. Oktober 2001. (Abhandlungen der Geistes- und sozialwissenschaftlichen Klasse Einzelveröffentlichung Nr. 9). Mainz: Akademie der Wissenschaften. Stuttgart: Franz Steiner Verlag
- Erber, Katharina: Auf dem Weg zur „digitalen Bibliothek“, In: ABI-Technik, 17 (1997), Heft 2, S. 143-153
- Hänel, Matthias: Vorwort. In: Bibliotheksforum Bayern, 28 (2000), Heft 2, S. 131. München: Sauer
- Hilz, Helmut: JSTOR – ein Projekt zur Zeitschriftendigitalisierung in den USA. In: Zeitschrift für Bibliothekswesen und Bibliographie, 46 (1999), Heft 3, S. 213-225
- Huck, Thomas; Striefler; Hubert-Günter: Retrospektive Digitalisierung von historischen Presseartikeln auf Papier, Rollfilm und Mikrofiche der Archive des Hamburgischen Welt-Wirtschafts-Archivs (HWWA). Zwischenbericht zum 15.12.2004 (unveröffentlicht)
- Klinge, Michael; Schüler, Mechthild: Das DFG-Projekt zur Digitalisierung der seltenen Bücher, Karten, Manuskripte zur Erforschung Sibiriens aus der Sammlung Asch. In: Bibliotheksdienst, 37 (2003), Heft 6, S. 740-753
- Leskien, Hermann: Retrodigitalisierung – Eine Zwischenbilanz. In: Bibliotheksforum Bayern, 28 (2000), Heft 2, S. 132-153. München: Sauer
- Mittler, Elmar: Verteilte digitale Forschungsbibliothek. Ein neuer Förderbereich der Deutschen Forschungsgemeinschaft. In: Wefers, Sabine (Hrsg.): 7. Deutscher Biblio-

theyskongress, 87. Deutscher Bibliothekartag in Dortmund 1997. Von Gutenberg zum Internet. S. 81-87

Mittler, Elmar (Hrsg.): Retrospektive Digitalisierung von Bibliotheksbeständen: Berichte der von der Deutschen Forschungsgemeinschaft einberufenen Facharbeitsgruppe „Inhalt“ und „Technik“; [DFG-Projekt Vorbereitung des Aufbaus einer verteilten digitalen Forschungsbibliothek in bibliothekarischer, fachlicher und technischer Hinsicht] / Deutsches Bibliotheksinstitut. [Projektkoordinator und Hrsg. Elmar Mittler]. – Berlin: Deutscher Bibliotheksdienst, 1998 (Dbi-Materialien ; 166 : Schriften der Deutschen Forschungsgemeinschaft)

Nordmann, Frauke; Striefler, Hubert-Günter: Antrag für die Förderung des Projekts „Retrospektive Digitalisierung von historischen Presseartikeln auf Papier, Rollfilm und Mikrofiche der Archive des Hamburgischen Welt-Wirtschafts-Archivs (HWWA)“ durch die Deutsche Forschungsgemeinschaft, Hamburg, 02.06.2003 (unveröffentlicht)

Ritzi, Christian: Pädagogische Zeitschriften und Nachschlagewerke als elektronische Dokumente im Internet. In: B.I.T. online, 3 (2000), Heft 1, S. 61-70

Schibel, Wolfgang; Kredel, Heinz: Vom Spezialbestand alter Drucke zum kommentierten Volltextkorpus im WWW. Das Projekt CAMENA (Neulateinische Dichtung Deutschlands). In: Buch, Thomas et al. (Hrsg.) (2001): Standards und Methoden der Volltextdigitalisierung. Beiträge des Internationalen Kolloquiums an der Universität Trier, 8./9. Oktober 2001. (Abhandlungen der Geistes- und sozialwissenschaftlichen Klasse Einzelveröffentlichung Nr. 9). Mainz: Akademie der Wissenschaften. Stuttgart: Franz Steiner Verlag

Wawra, Steffen: Eine virtuelle Akademiebibliothek? Mehrwert durch Digitalisierung und Metadaten. In: Schmidt, Ralph (Hrsg.): DGI-Proceedings '99 - '01, S. 211-222

#### Anmerkung zu Internet-Quellen:

Wurde direkt von Internet-Seiten zitiert, sind URL und Zugriffsdatum direkt am Ende des Zitats genannt. Handelt es sich um Hinweise auf umfassende Internetangebote wie die Homepages beispielhaft genannter Retrodigitalisierungsprojekte, so befinden sich die URLs im Fußnotentext.

## **6 Anhang**

### **Anhang 1: Liste von der Deutschen Forschungsgemeinschaft geförderten Projekte im Förderprogramm „Retrospektive Digitalisierung von Bibliotheksbeständen“**

Entnommen dem Internetangebot des Göttinger Digitalisierungszentrums (<http://dgz.sub.uni-goettingen.de/de/vdf-d/vdf-liste.shtml>) sowie sämtliche auf der nächsten unteren hierarchischen Ebene liegende Dokumente)

### **Anhang 2: Graphische Darstellung der Infrastruktur zur Erschließung und Bereitstellung der digitalisierten Dokumente**

(eigener Entwurf)

### **Anhang 3: Entwurf der Homepage für die Präsentation der digitalisierten Zeitungsausschnitte des Personenarchivs**

(<http://webopac2/digiview/DIGI.cfm>; Stand: 07.02.2005)



Hiermit erkläre ich, dass ich die vorliegende Abschlussarbeit selbständig und ohne fremde Hilfe angefertigt und alle Stellen, die wörtlich oder annähernd wörtlich aus Veröffentlichungen entnommen sind, als solche kenntlich gemacht habe und mich keiner anderen als der angegebenen Literatur und Hilfsmittel bedient zu haben.

Hamburg, 20.02.2005