

# Bayerische Staatsbibliothek

## HILDEGARD SCHÄFFLER / ASTRID SCHOGER / MARGARETE WITKЕ

### Langzeitarchivierung an einer Universalbibliothek: Praxis und Perspektiven aus Sicht der Bayerischen Staatsbibliothek

Digitale Publikationen nehmen im Wissenschaftsbetrieb wie auch im gesellschaftlichen Leben insgesamt einen immer höheren Stellenwert ein. Die Erhaltung des digitalen kulturellen und wissenschaftlichen Erbes stellt nicht zuletzt die Bibliotheken angesichts des raschen technologischen Wandels im Bereich der Hard- und Software vor eine neue Aufgabe: die Langzeitarchivierung (LZA) zur Sicherung der dauerhaften Verfügbarkeit ihrer erworbenen, gesammelten und selbst produzierten digitalen Ressourcen.

Im Folgenden wird zunächst skizziert, in welcher Weise die BSB in überregionale und nationale Aktivitäten zur Langzeitarchivierung als einer nur gemeinschaftlich zu lösenden Aufgabe eingebunden ist. Sodann wird ausführlich die bereits bewährte Praxis bei der Archivierung der eigenen Digitalisate beschrieben. Mit einem soeben begonnenen Pilotprojekt zur Langzeitarchivierung heterogener Netzpublikationen verschiedener Provenienz sollen diese Erfahrungen auf das gesamte Spektrum ausgedehnt werden. Abschließend werden die Bemühungen und mögliche Perspektiven der Bayerischen Staatsbibliothek in Zusammenhang mit dem Erhalt von datenträgerbundenen Medien dargestellt.

Digital publications are acquiring an ever larger role in the academic environment as well as for the general public. In light of the rapid technological transformations in the area of hardware and software, maintaining the society's digital cultural and scientific heritage means that libraries, among others, are faced with a new responsibility: Long-term preservation (in German: LZA) for ensuring the permanent accessibility of acquired, collected and self-produced digital resources. This article outlines the ways in which the Bavarian State Library (BSB) is linked with regional and national programs for long-term preservation – a task only possible as a collective undertaking. This is followed by a detailed description of the tried and proven practice of archiving the library's own digitizations. A newly initiated pilot project for long-term preservation will now extend previous experience in this area to the wide-range spectrum of digital holdings. Finally, the article presents the BSB's efforts and potential future steps toward preserving data-storage media.

#### LANGZEITARCHIVIERUNG ALS AUFGABENFELD EINER UNIVERSAL- UND ARCHIVBIBLIOTHEK

Die Bayerische Staatsbibliothek (BSB) ist eine wissenschaftliche Universalbibliothek mit einem lokalen, regionalen und überregionalen Aufgabenspektrum. Neben der Literaturversorgung am Ort fungiert die BSB auch als zentrale bayerische Landes- und Archivbibliothek mit Pflichtexemplarrecht für Bayern. Auf überregionaler Ebene nimmt die BSB mit einer Reihe von geisteswissenschaftlichen Fächern am Sonder-sammelgebietsplan der Deutschen Forschungsgemeinschaft (DFG) teil. Der bedeutende Altbestand sowie eine Reihe von Sondersammlungen machen die BSB nicht zuletzt auch zu einer internationalen Forschungsbibliothek.

Die Vielfalt der Aufgaben auf den unterschiedlichsten Ebenen findet ihre Fortsetzung im Bereich der digi-

talen Medien, sei es auf physischen Datenträgern oder als Netzpublikationen, für die sich analog zum konventionellen Bestand ein Archivierungsauftrag ableiten lässt. Zu nennen sind hier insbesondere die nachfolgend skizzierten Felder:

#### Lizenpflichtige E-Medien

Abgesehen von der Versorgung der lokalen Klientel mit Zugriffsmöglichkeiten auf lizenpflichtige digitale Medien im Rahmen des Erwerbungs- und Sammelprofils der Bibliothek wurde die BSB von den bayerischen Hochschulbibliotheken damit beauftragt, in enger Abstimmung mit den Teilnehmern Konsortialabkommen über elektronische Zeitschriften und Datenbanken zu verhandeln und abzuschließen.<sup>1</sup> Im Bereich der Netzpublikationen werden sowohl für die lokal lizenzierten Medien als auch für die im Rahmen des Bayern-Konsortiums erworbenen digitalen Materialien nach Möglichkeit Vereinbarungen über die Langzeitverfügbarkeit der lizenzierten Periodika und Datenbanken getroffen. Da die von den Verlagen angebotenen Archivierungsoptionen nicht unbedingt langfristige Sicherheit bieten, müssen Möglichkeiten geschaffen werden, die erworbenen Daten auch selbst zu speichern und im Rahmen der vorliegenden Rechte zur Benutzung zur Verfügung zu stellen. Für digitale Medien auf Datenträgern stellt sich die Frage der dauerhaften Benutzbarkeit der Ressourcen vor dem Hintergrund der Haltbarkeit des Trägermediums bzw. der notwendigen Hard- und Softwarekompatibilität künftiger Geräte und Betriebssysteme. Ein langfristiger Erfolg kann nur durch Lösung der Informationen vom Datenträger erzielt werden.

#### Pflichtablieferung und Amtliche Veröffentlichungen

Als regionale Pflichtexemplarbibliothek für Bayern erhält die BSB im Rahmen des geltenden Landesgesetzes über die Pflichtablieferung bereits jetzt elektronische Medien auf physischen Datenträgern, was einen entsprechenden Archivierungsauftrag impliziert. Für den Bereich der Netzpublikationen sind gesetzliche Regelungen in Vorbereitung (vgl. dazu auch unten, Der überregionale und nationale Rahmen). Einschlägig ist darüber hinaus in diesem Zusammenhang aber auch der Bereich der frei im Internet zugänglichen *Bavaria*.<sup>2</sup> Über die Pflichtablieferung im engeren Sinne hinaus hat die BSB auch bei Amtlichen Veröffentlichungen von Regional- und Bundesbehörden sowie Gebietskörperschaften einen Sammelauftrag. Publikatio-



Hildegard Schäffler

Foto privat



Astrid Schoger

Foto privat



Margarete Wittke

Foto privat

gesetzliche Regelungen  
in Vorbereitung

nen dieser Art stellen zunehmend das Erscheinen in gedruckter Form ein. Eine Verlinkung auf die einschlägigen Webseiten führt zwar kurzfristig zur gewünschten Information, die Langzeitsicherung durch die herausgebenden Stellen ist aber keinesfalls gewährleistet. Auch hier besteht ein möglichst mit anderen Sammelstellen abgestimmter Handlungsbedarf auf Seiten der Archivbibliothek.

#### Eigene Retrodigitalisate

Im Jahr 1997 wurde mit Unterstützung der Deutschen Forschungsgemeinschaft das Münchener Digitalisierungszentrum eingerichtet, das mittlerweile als Sachgebiet Digitalisierung im Referat Digitale Bibliothek der Abteilung Bestandsaufbau und Erschließung verstetigt wurde. Im Rahmen der Konzeption des »retrodigitalen Bestandsaufbaus« wurde 2004 mit Hilfe der Förderung des Bayerischen Staatsministeriums für Wissenschaft, Forschung und Kunst eine »Digitalisierungsstraße« eingerichtet. Damit steht nun eine Infrastruktur zur systematischen Eigendigitalisierung von Druckwerken und Sondermaterialien wie z.B. Handschriften und Karten auch im Rahmen von Kooperationsprojekten oder als Digitization on Demand zur Verfügung. Die auf diese Weise entstehenden großen Datenmengen bedürfen ebenfalls der Langzeitsicherung.

#### Elektronische Medien in den Sondersammelgebieten

Die überregionale Literaturversorgung im Rahmen der SSG-Verpflichtung – die BSB pflegt im Rahmen des Sondersammelgebietsplans der DFG die Fächer Geschichte, Osteuropa, Klassische Altertumswissenschaften und Musikwissenschaft – steht mit der Einführung elektronischer Medien vor neuen Herausforderungen. Dies bedeutet zum einen, dass wissenschaftlich relevante Internetquellen systematisch gesammelt und erschlossen werden.<sup>3</sup> Zumindest für einen Teil dieser Quellen, deren dauerhafte Zugänglichkeit auf den Seiten der Urheber nicht gesichert ist, sieht sich die BSB als zuständige SSG-Bibliothek im Einvernehmen mit den Rechteinhabern in der Archivierungspflicht. Bezogen auf lizenzpflichtige Medien sind an der BSB mehrere Pilotprojekte angesiedelt, die sich mit der Frage der überregionalen Literaturversorgung mit zugriffsbeschränkten E-Ressourcen befassen. Nach dem DFG geförderten Projekt »Elektronische Zeitschriften in der überregionalen Literaturversorgung« (EZUL), in dem Geschäfts- bzw. Nutzungsmodelle für den kontrollierten überregionalen Zugang zu SSG-relevanten elektronischen Zeitschriften entwickelt wurden,<sup>4</sup> wird derzeit im Rahmen eines weiteren DFG-Projekts auf der Grundlage entsprechender Vertragsverhandlungen eine Infrastruktur für den Pay-per-Use-Zugang zu SSG-relevanten Datenbanken aufgebaut.<sup>5</sup> Komple-

mentär hierzu hat die Deutsche Forschungsgemeinschaft zu Beginn des Jahres 2005 im Rahmen einer Sonderfördermaßnahme den dauerhaften Erwerb umfangreicher digitaler Textsammlungen auf nationaler Ebene finanziert. Die BSB fungierte hier neben der SUB Göttingen, der UB Frankfurt und der SBB-PK Berlin als Verhandlungsführerin. Aus beiden Ansätzen – Pay-per-Use und Nationallizenzen – ergibt sich die Notwendigkeit der dauerhaften Sicherung der erworbenen Daten. Im Hinblick auf Nationallizenzen-Datenbanken kann diese Aufgabenstellung bereits in nächster Zeit angegangen werden, während für Lösungsansätze im Pay-per-Use-Bereich eher von einer mittel- und langfristigen Perspektive auszugehen ist.

#### DER ÜBERREGIONALE UND NATIONALE RAHMEN: DIE ARBEITSGEMEINSCHAFT REGIONALBIBLIOTHEKEN UND DAS KOMPETENZNETZWERK NESTOR

Die Auseinandersetzung mit dem Thema Langzeitarchivierung digitaler Ressourcen an der Bayerischen Staatsbibliothek ist eingebettet in überregionale und nationale Aktivitäten. Dazu zählt zum einen die Mitarbeit in einer Arbeitsgruppe der Arbeitsgemeinschaft Regionalbibliotheken im Deutschen Bibliotheksverband Sektion 4 (AG Regionalbibliotheken), die sich aus Sicht der regionalen Pflichtexemplarbibliotheken mit der Novellierung der Gesetzgebung zur Pflichtablieferung befasst, und zum anderen die Beteiligung am Gemeinschaftsprojekt nestor, das ein Kompetenznetzwerk für die Langzeitarchivierung und Langzeitverfügbarkeit digitaler Quellen für Deutschland aufbaut.

Die AG Regionalbibliotheken hat einen Musterentwurf zu Landesgesetzen über die Sammlung von Pflichtexemplaren erarbeitet, der durch eine gemeinsam mit Der Deutschen Bibliothek verfasste Stellungnahme ergänzt wird. Die KMK-AG der Bibliotheken hat Ende 2004 auf der Basis dieses Musterentwurfs eine Empfehlung an die Länder zur Umsetzung in der Gesetzgebung ausgesprochen. Es ist davon auszugehen, dass die Ländergesetzgebung der noch für das Jahr 2005 erwarteten Novellierung auf Bundesebene folgen wird.<sup>6</sup> Neben der skizzierten Vorbereitung eines gesetzlichen Rahmens für die Novellierung der Pflichtexemplargesetzgebung hat eine Unter-AG einen Katalog von möglichen Auswahlkriterien für das Sammeln von Netzpublikationen erarbeitet.<sup>7</sup> In diesem Zusammenhang wurde eine Typologie der E-Ressourcen entwickelt, aus der sich Kriterien für die Sammelwürdigkeit ableiten lassen. Diese Überlegungen zum Auswahlprofil bauen letztlich auf der Einsicht auf, dass Netzpublikationen nicht in der Vollständigkeit gesammelt und damit auch archiviert werden können, wie

dies für den konventionellen Bereich zumindest angestrebt wird.

Das Gemeinschaftsprojekt nestor,<sup>8</sup> das die Bayerische Staatsbibliothek zusammen mit anderen Gedächtnisorganisationen initiiert hat, ist aus der Überzeugung heraus entstanden, dass der Erhalt des kulturellen Erbes nur gemeinschaftlich gelingen kann. Dieses durch das BMBF geförderte Projekt hat das Ziel, ein Kompetenznetzwerk zur Langzeitarchivierung aufzubauen. Aufgaben des Projektes sind unter anderem: das Bewusstsein für die Probleme der LZA in Deutschland zu schaffen, das Wissen und die Kompetenzen in diesem Bereich zu mehren und zu verbreiten, eine kooperative Lösung für Deutschland vorzubereiten und in den internationalen Kontext einzubinden. Neben der Beteiligung an allen gemeinschaftlichen Aufgaben liegt der inhaltliche Schwerpunkt der Mitarbeit der BSB auf der Entwicklung eines Kriterienkatalogs für vertrauenswürdige digitale Langzeitarchive. Um die Vertrauenswürdigkeit der Archive nachzuweisen, sollen jene Aspekte ihres organisatorischen Aufbaus und der verwendeten technischen Verfahren geprüft und bewertet werden, die für die Langzeitarchivierung relevant sind. Die Kriterien und deren Bewertungen dienen als Grundlage für die Entwicklung eines Zertifizierungsverfahrens digitaler Langzeitarchive. Zusammen mit der Humboldt-Universität zu Berlin hat die BSB die Federführung der nestor-Arbeitsgruppe »Vertrauenswürdige Archive – Zertifizierung« übernommen, die sich dieses Themas annimmt und Experten aus allen Lebenszyklen digitaler Objekte von der Produktion über die Bereitstellung und Nutzung bis zur Archivierung einbezieht. Dabei baut sie auf Vorarbeiten der international besetzten RLG/OCLC Working Group on Digital Archive Attributes<sup>9</sup> auf, an der auch die BSB beteiligt war.<sup>10</sup>

#### VON DER DIGITALISIERUNG ZUR LANGZEITARCHIVIERUNG: DIE PRAXIS AN DER BAYERISCHEN STAATSBIBLIOTHEK

Seit dem Beginn der Digitalisierung 1997 hat die BSB ca. 2,4 Terabyte digitaler Masterdaten erstellt und in Zusammenarbeit mit dem Münchener Leibniz-Rechenzentrum archiviert. Diese Daten wurden durch Drittmittelprojekte (z.B. im Rahmen des DFG-Förderprogramms »Retrospektive Digitalisierung von Bibliotheksbeständen« sowie »Bayerische Landesbibliothek Online«), durch »Digitization on Demand« und in Eigenleistung finanziert. Zukünftig sind hohe Zuwachsrate an digitalen Daten zu erwarten, da mit der systematischen Digitalisierung (»retrodigitaler Bestandsaufbau«) begonnen wurde.<sup>11</sup> Alle im Rah-

men der Retrodigitalisierung erstellten Daten stehen als frei verfügbare Publikationen im WWW bereit.<sup>12</sup> Die ersten konkreten Schritte in Richtung Langzeitarchivierung wurden an der BSB im Zusammenhang mit den Retrodigitalisaten unternommen, weshalb im Folgenden der auf diese Weise entstandene Workflow exemplarisch dargestellt werden soll.

Das Sachgebiet Digitalisierung hat sich seit dem ersten Retrodigitalisierungsprojekt, den Reichstagsprotokollen von 1867 bis 1895, an Verfahren guter Praxis orientiert und diese zum Teil mitgestaltet.<sup>13</sup> In der konsequenten Anwendung offener und offen gelegter Standards für die Bilddigitalisierung und Erschließung liegt die Voraussetzung für eine langfristige Archivierung des Materials und die Migration in unterschiedliche Bereitstellungssysteme, die für einen Teil der Daten bereits erfolgreich durchgeführt wurde.

Der digitale Master entsteht immer im TIF-Format (Tagged Image File Format), das höchste Qualität und verlustfreie Speicherung der Bilddaten garantiert. Von diesem Master können alle möglichen Derivate wie JPGs, GIFs oder PDFs gebildet werden, die für die Darstellung im Internet notwendig sind. Die Herstellung von Master-Images und deren Archivierung auf verteilten Datenträgern bzw. -systemen kann unersetzblichen Wissensbestand vor unwiderruflichem Verlust – etwa durch Brand – schützen. Ausbelichtungen auf Filmmaterial zur technikunabhängigen Sicherung sind auf der Basis eines digitalen Masters ebenso möglich wie der Druck von hochwertigen Faksimiles. Die Retrodigitalisierung dient somit nicht nur dem Schutz vor Verlust, sondern auch der Schonung und dem langfristigen Erhalt der Originale.<sup>14</sup> Nach der Bereitstellung im Internet müssen diese in der Regel nicht mehr für weitere Benutzungen ausgehoben werden. Sie stehen der Forschung weltweit zur Verfügung.

Die Struktur der entstehenden digitalen Objekte wird in einer XML-Datei beschrieben. Die BSB bezieht sich beim Erstellen der XML-Dateien auf die Richtlinien und Standards der Text Encoding Initiative (TEI).<sup>15</sup> Die XML-Datei enthält im sogenannten TEI-Header die bibliografischen Daten des Objekts. Bei Bedarf kann die XML-Datei die tiefere Erschließung des Buches, etwa das Inhaltsverzeichnis, und gegebenenfalls den Volltext in maschinenlesbarer Form aufnehmen. In der einfachsten Form enthält sie neben den bibliografischen Angaben nur die Bezeichnung der Images in der Digitalisierungsreihenfolge. Dies ist die Voraussetzung dafür, dass die Digitalisate zu einem elektronischen Buch zusammengefügt werden können und eine Navigation auf dem Bildschirm möglich wird. Durch die Wahl des XML-Formates und des TEI-Standards sind diese Dateien ebenfalls für die Archivierung

#### Aufgaben des Kompetenznetzwerks nestor

#### digitale Master

#### XML-Dateien nach TEI-Standards

und Migration geeignet. Sie werden mit den Images gemeinsam und gleichzeitig archiviert.

Mittlerweile wurde innerhalb der BSB ein System entwickelt, das einerseits den Workflow unterstützt und andererseits die bibliografischen, strukturellen, technischen und administrativen Metadaten verwaltet. Es konnten damit sowohl im Prozess der Auftragsvergabe, der eigentlichen Digitalisierung und Erschließung als auch bei der Bereitstellung der digitalen Bücher im Netz und der Langzeitarchivierung eine Vielzahl von ehedem zeitraubenden Arbeitsschritten datenbankgestützt automatisiert werden. Somit wird die Vollständigkeit, Fehlerfreiheit und Rekonstruierbarkeit der Prozesse garantiert. Ferner ist dadurch die gleich bleibende hohe Qualität der Daten gewährleistet.

Die bis 2004 übliche Praxis, die digitalen Master auf CD zu speichern und blockweise an das Münchener Leibniz-Rechenzentrum zur Endarchivierung zu überspielen, wurde inzwischen zugunsten von Mehrfachspeicherungen auf räumlich verteilten Serversystemen aufgegeben. Dies kann allerdings nur dann verlässlich funktionieren, wenn auf allen Systemen eine vollständig einheitliche, eindeutige und unveränderliche Verzeichnisstruktur vorhanden ist. Erreicht wird dies durch datenbankgestützten und automatisierten Verzeichnisaufbau und Dateinamenvergabe.

Den Kern des Workflows, der nun die Herstellung des Digitalats von der Auftragsvergabe bis zur Archivierung an der BSB begleitet, bildet eine für jede bibliografische Einheit bereits bei der Auftragsvergabe in der »Zentralen Erfassungs- und Nachweisdatenbank für Digitalisate« (ZEND) erzeugte ID-Nummer. Sie ist unveränderlich in der Datenbank verankert und bestimmt sowohl den festen Stamm der Bezeichnung eines jeden Images als auch den Namen des Verzeichnisses, in welchem die Digitalate eines Bandes und die XML-Strukturinformation zusammengefasst werden. Sie stellt zudem die Verbindung zu den Daten-

bankrelationen dar, in denen Digitalisierungsaufträge mit den Angaben für die Ausführung und die technischen Metadaten verwaltet werden. Sie steuert darüber hinaus die Programme, die die automatisierte Anfertigung von browserkompatiblen Derivaten (JPG, GIF und PDF) durchführen, sowie die Tools für die Strukturdatenbearbeitung und die Bildnachbearbeitung. Schließlich bildet sie den buchspezifischen Kern der NBN (National Bibliography Number), eines persistenten Identifikators, der über Die Deutsche Bibliothek in Frankfurt verwaltet wird und die dauerhafte Ansprechbarkeit des Titels im Internet garantiert.<sup>16</sup> Im Folgenden wird der Workflow in einzelnen Schritten dargestellt.

#### Schritt 1: Metadaten und Identifikatoren in der Zentralen Nachweis- und Erfassungsdatenbank (ZEND).

Die ZEND ist über eine Z39.50-Schnittstelle mit dem lokalen Katalogsystem der Bayerischen Staatsbibliothek verbunden. Durch Einkopieren der Katalognummer in ein Online-Formular werden die bibliografischen Daten des zu digitalisierenden Titels in verkürzter Form in die Datenbank übernommen. Mit der Erfassung eines Titels in ZEND sind unmittelbar alle endgültigen und unveränderlichen Identifikatoren vergeben. Über ein elektronisches Auftragsformular wird der Digitalisierungsauftrag erzeugt. Beim Absenden wird überprüft, ob die eingegebenen Parameter plausibel sind, und die technischen und administrativen Metadaten werden abgespeichert. Zu diesen gehören etwa Angaben zur Objektart (z.B. Karte, Zeitschrift), zur geforderten Bildauflösung und Farbtiefe oder besondere Hinweise für die Einstellung der Scanner, wie sie beispielsweise beim Digitalisieren von gold- oder silberfarbigen Abbildungen notwendig sind. Mitgeführt werden darüber hinaus das Bestelldatum, der Name und die Abteilung des Auftraggebers sowie der Ausführenden bzw. die Anschrift des externen Dienstleisters.

Abb. 1:  
Digitalisierungsauftrag  
mit Identifikatoren und  
Digitalisierungsparametern

Bayerische Staatsbibliothek, Ludwigstr. 16, 80539 München - Abt. EA-ES-DB - Wittke
<b>Digitalisierungsauftrag (Fotostelle) / Projekt:</b>
<b>Signatur :4 Bor. 24 i</b> Gregorius, Immanuel F.: Historische Nachricht von dem Pfarrkirchthurm, und den vorzeiten und jetzt darauf befindlichen Glocken zu Lauben. [1775] <b>urn:nbn:de:bvb:12-bsb00001084-3</b>
<b>Digitalisierungsparameter:</b>
Auflösung: 400 - Farbe
Dateiname der ersten Seite: <b>bsb00001084_00001.tif</b>
Anzahl der Seiten pro Image: <b>1</b>
Nachbearbeitung: ---
Besonderes: -----

## Schritt 2: Digitalisierung

Die Digitalisierung wird hausintern oder vom externen Dienstleister nach den im Auftrag vorgegebenen Parametern durchgeführt. Nach dem Digitalisieren werden die Images ohne Bildung von Zwischenverzeichnissen in einem Sammelverzeichnis auf dem FTP-Server der BSB abgelegt.

## Schritt 3: Automatisierte Erzeugung der Strukturdatei und Bereitstellung

Das Sammelverzeichnis, in welches die Images abgelegt werden, wird mit Hilfe von zeitgesteuerten Verfahren überwacht. Immer dann, wenn Bilddaten zur Weiterverarbeitung dort vorhanden sind, wird das Bereitstellungsverfahren gestartet. Zunächst wird die Datenbank anhand der Dateinamenbezeichnung abgefragt, ob Titeldaten erfasst sind. Sofern eine Konkordanz vorhanden ist, wird automatisch im Bereitstellungsbereich auf dem öffentlich zugänglichen Server ein Verzeichnis unter dem Namen der ID des Digitalisats eingerichtet. In dieses Verzeichnis werden alle zusammengehörenden Images kopiert. Die Dateien werden geordnet und ihre Namen in der richtigen Reihenfolge in die XML-Strukturdatei geschrieben. Über die Datenbankabfrage werden die Titelangaben ausgelesen und im TEI-Header-Format ebenfalls in die XML-Datei eingefügt. Schließlich werden aus den TIFF-Dateien JPG-Derivate hergestellt. Dafür wird der TIFF-Header mit den Angaben zu Größe, Auflösung, Ausrichtung etc. ausgelesen. Die Image-Größe für das im Web später sichtbare JPG-Image wird in Relation zum Originalbild festgelegt, wobei die Images auf ca. 20 % der Originalgröße verkleinert werden. In der Regel reicht dies für gute Lesbarkeit aus. Um auch Digitalisate in sehr kleiner Schrift lesbar anzubieten, wird zusätzlich ein JPG-Image in der Größe von ca. 45 % des Originals angelegt. Nach dem Schreiben der XML-Struktur-Datei und der Bildherstellung für das Web ist eine einfach zu durchblätternde digitale Ausgabe des Buches fertig und benutzbar. Vom Ablegen der digitalen Master im Sammelverzeichnis bis zur Fertigstellung dieser einfachen Webversion sind keine manuellen Tätigkeiten erforderlich. Für die Produktion eines 200 Seiten umfassenden Buches benötigt ein leistungsstarker Rechner zurzeit etwa eine Stunde.

## Schritt 4: Qualitätskontrolle und Nachbearbeitung

Die fertig gestellte digitale Ausgabe eines Buches kann jederzeit tiefer erschlossen und gegebenenfalls auch korrigiert werden. Dafür stehen eine ganze Reihe von Tools zur Verfügung, wovon als wichtigster der XML-Editor zu nennen ist, der von jedem Ort innerhalb der Staatsbibliothek webgestützt aufrufbar ist und es

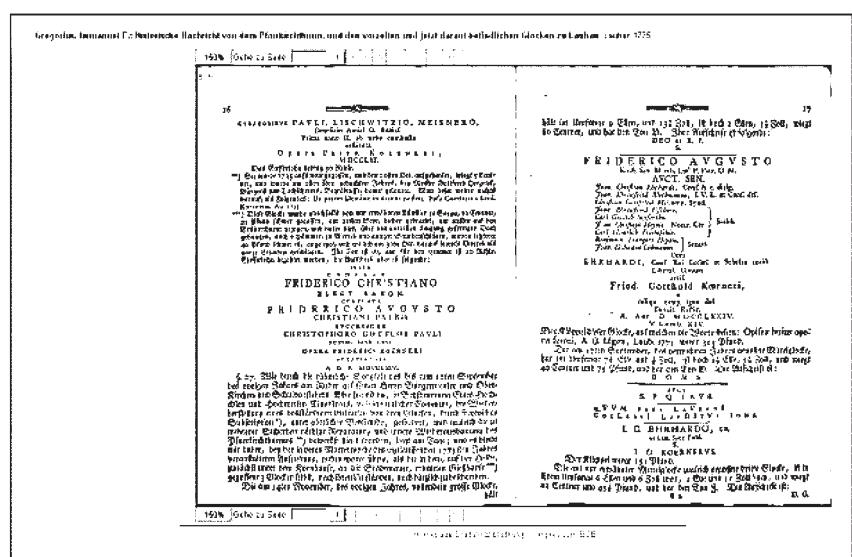


Abb. 2: Einfache Webversion mit Blätterfunktionen

erlaubt, Kapitelüberschriften zuzuweisen und Seitenzählungen (arabisch, römisch, blatt-, spalten- und seitenseitweise) halbautomatisch einzufügen. Komplexere Zählungen können auch per Hand bei jeder Seite eingefügt werden. Vom XML-Editor aus lässt sich auch eine Bildkorrektur – etwa durch Drehen einzelner Seiten – anstoßen. Eine tiefere Erschließung des Inhalts wird insbesondere bei den großen Digitalisierungsprojekten wie der Digitalisierung ganzer Zeitschriften oder vielbändiger Werke durchgeführt. Hier werden Aufsatztitel, Register, Lemmata etc. erfasst und vielfach auch der Volltext über OCR gewonnen.

## automatische Weiterverarbeitung des digitalen Masters



Abb. 3: XML-Editor mit Bildbearbeitungsfunktionen

## Schritt 5: NBN-Vergabe und Archivierung

Mit der Freigabe für die Benutzung im Internet werden NBN und aktuelle URL über eine OAI-Schnittstelle der ZENTD dem Frankfurter NBN-Management übermittelt, das dafür sorgt, dass die dem Digitalisat zugewiesene NBN im Internet aufrufbar wird. Die NBN

#### **testweise Archivierung heterogener Netzpublikationen**

wird auf Verbundebene und im lokalen Katalogsystem der BSB nachgewiesen, so dass Benutzer, die den Titel recherchieren, direkt auf das Digitalisat hingewiesen werden. Zeitgleich mit der Freigabe des Objekts wird die Archivierung am Münchener Leibniz-Rechenzentrum aktiviert. Dafür ist in der BSB ein Client des Archiv- und Backup-Systems Tivoli Storage Manager von IBM installiert, dem in einer allnächtlich laufenden Prozedur mitgeteilt wird, welche bereits zur Benutzung freigegebenen Objekte noch nicht der Langzeitarchivierung zugeführt worden sind. Auf diese greift der Tivoli Storage Manager gezielt zu und bildet einen Bereich im Archivierungssystem, der nach der ID bezeichnet wird. Hier werden alle zum Objekt gehörigen Daten abgespeichert, wobei die Informationen über den Verzeichnisaufbau auf dem Bereitstellungsserver ebenfalls archiviert werden. Damit können die archivierten Dateien innerhalb kürzester Zeit zurückkopiert werden. Für die Langzeitarchivierung kommt dabei den XML- und TIFF-Dateien die größte Bedeutung zu. Die zusätzliche Archivierung der Bildderivate bringt weitere Sicherheiten, insbesondere für den Fall eines Defektes des öffentlich zugänglichen Bereitstellungsservers. Im Vergleich zum Rückkopieren wäre die Neuanfertigung von Hunderttausenden von JPG-Images zweifellos ein langwieriger Prozess. Nach dem Archivieren eines Objekts mit allen seinen Einzeldateien erfolgt die Rückmeldung an die Datenbank, in der dieser Arbeitsschritt wie alle anderen auch in den administrativen Metadaten dokumentiert wird. Das Leibniz-Rechenzentrum selbst speichert die zu archivierenden Daten mehrfach auf verschiedenen räumlich getrennten Speichermedien.

#### **Zusammenfassung der datenbankgestützten Arbeitsprozesse**

Zum Workflow der Digitalisierung und Archivierung an der BSB lässt sich zusammenfassend festhalten, dass die Verwendung von offenen und offen gelegten Standards für die Images und die Beschreibungsdaten (Datenformate TIFF, XML/TEI) einerseits die Langlebigkeit der durch Digitalisierung erzeugten Daten sicherstellt, andererseits die Migration zu neuen Beleistungssystemen ermöglicht. Die Archivierung garantiert den Erhalt dieser Daten, die verwendeten persistenten Identifikatoren den dauerhaften Zugriff. Die Automatisierung aller Arbeitsschritte von der Auftragsvergabe bis zur Archivierung sichert zudem die hohe Qualität der Objekte. Die Verwaltung bibliografischer, struktureller, technischer sowie administrativer Metadaten ermöglicht zukünftig die gezielte Anwendung von Langzeiterhaltungsstrategien. Damit ist die BSB für die Anforderungen der Zukunft gerüstet.

#### **LANGZEITARCHIVIERUNG VON HETEROGENEN NETZPUBLIKATIONEN – EIN PILOTPROJEKT**

Wie oben dargestellt, trägt die Bayerische Staatsbibliothek als Universal- und Landesbibliothek Verantwortung für ein breites Spektrum an digitalen Objekten. Deshalb beabsichtigt sie, das skizzierte Vorgehen bei der Archivierung der eigenen Digitalisate nun um konkrete Erfahrungen mit der Langzeitarchivierung von Netzpublikationen – von der Sammlung über die Verwaltung und Speicherung bis hin zur Nutzung – aus dem gesamten heterogenen Spektrum zu ergänzen. Zusammen mit dem Leibniz-Rechenzentrum, das bereits an der Archivierung der eigenen Digitalisate beteiligt ist, hat sie vor kurzem mit einem durch die DFG geförderten Pilotprojekt zur Langzeitarchivierung von Netzpublikationen begonnen. Um die Medienvielfalt der BSB möglichst umfassend zu repräsentieren, sollen Medien verschiedener Provenienz in das Projekt einbezogen werden. Darunter fallen:

- Veröffentlichungen von potenziellen Pflichtverlagen,
- im Bayernkonsortium lizenzierte E-Zeitschriften und Datenbanken,
- amtliche Veröffentlichungen von Regional- und Bundesbehörden sowie Gebietskörperschaften,
- wissenschaftlich relevante Literatur aus der nationalen und internationalen Produktion in den SSG-Fächern Geschichte, Altertumswissenschaften, Osteuropa und Musikwissenschaft,
- in den Fachportalen zusammengestellte und erschlossene Netzpublikationen,

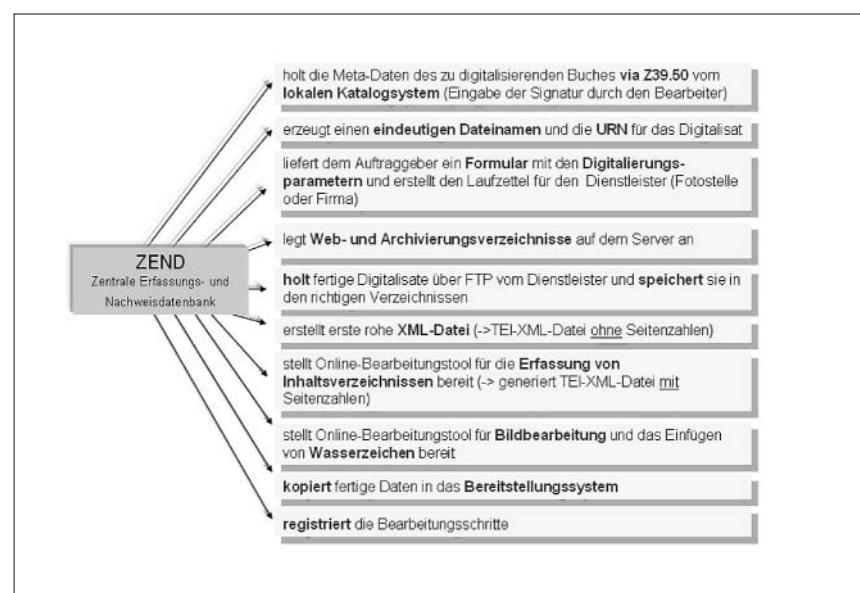


Abb. 4: Zusammenfassung der datenbankgestützten Arbeitsprozesse

- eigene und in Kooperationsprojekten erstellte Retrodigitalisate.

Das Datenmanagement von der Sammlung über die Erschließung bis zur Nutzung wird in der Bibliothek angesiedelt sein und in die bestehenden Bibliothekssysteme integriert werden. Ferner wird am Leibniz-Rechenzentrum das Archiv- und Backupsystem Tivoli Storage Manager als Archivspeichersystem zum Einsatz kommen, das sich bereits für die Archivierung der Digitalisate bewährt hat.

Solange grundlegende rechtliche Fragen bei der Langzeitarchivierung (z.B. die Pflichtablieferung, das Anfertigen mehrfacher Kopien, verändernde Eingriffe bei der Konvertierung bzw. Migration, Bereitstellung über Netze) noch nicht allgemein geregelt sind, können nur in Absprache mit den Produzenten die einzelnen Arbeitsschritte bei der Langzeitarchivierung modelliert und realisiert werden. Darin liegt einer der Schwerpunkte dieses Projekts. Erste Gespräche mit den Produzenten – Verlage, Regional- und Bundesbehörden, Wissenschaftler, Bibliotheken – wurden bereits geführt. Ferner wurde auch damit begonnen, Genehmigungen für die Archivierung von Websites einzuholen. Speziell in Bezug auf urheberrechtlich geschützte digitale Medien ist mit den Anbietern bzw. Produzenten weiterführend die Frage nach möglichen Geschäfts- und Nutzungsmodellen zu diskutieren. Die BSB hat in Bezug auf lizenpflichtige E-Medien ein dreistufiges Zugriffskonzept entwickelt, das von der lokalen Nutzung in den Räumen der Bibliothek über den externen Zugriff für registrierte BSB-Benutzer bis hin zu speziell vereinbarten Pay-per-Use-Modellen für die überregionale Literaturversorgung reicht. In welcher Weise sich diese Modelle etwa auf den Bereich der Pflichtablieferung digitaler Medien übertragen lassen bzw. welche Art von Modifikation erforderlich ist, wird im Projekt zu prüfen sein.

Die Entwicklung neuer bzw. die weitere Ausgestaltung existierender Workflows in der Bibliothek sowie im Rechenzentrum wird einen weiteren Schwerpunkt darstellen. Die enge Verzahnung mit den bewährten Arbeitsabläufen garantiert den nahtlosen Übergang zur Routine und damit die Nachhaltigkeit der hier zu entwickelnden Modelle.

Vorgesehen ist ferner die Erprobung von Methoden der Langzeitarchivierung, die sowohl den physischen Erhalt der Daten als auch den Erhalt der Verfügbarkeit betreffen. Wegen des geplanten Umzugs des Leibniz-Rechenzentrums und der Einführung neuer Hard- und Software während der Projektlaufzeit wird eine echte Migration des Archivspeichersystems stattfinden. Darüber hinaus wird die Migration ausgewählter Datenformate getestet werden.

Die Ergebnisse aus diesem Projekt, insbesondere die gemeinsam mit den Produzenten entwickelten Konzepte, werden unmittelbar in die gesamtdeutsche Diskussion im Rahmen von nestor und der AG Regionalbibliotheken eingehen und damit einen Beitrag zur gemeinsamen Lösung der Langzeitarchivierungsfrage in Deutschland leisten.

### **LANGZEITARCHIVIERUNG VON DATENTRÄGERGEBUNDENEN MEDIEN – PERSPEKTIVEN**

Die Bayerische Staatsbibliothek hat zusammen mit dem Institut für Softwaretechnologie der Universität der Bundeswehr München bereits in den Jahren 1999–2001 in einem sehr frühen von der DFG geförderten Projekt zur »Langzeitarchivierung digitaler Medien« vorbereitende Schritte in Richtung LZA unternommen. Zunächst wurde eine Erhebung über den Typ, die Anzahl, den Inhalt sowie die Gefährdung von digitalen Publikationen auf Datenträgern im Bestand der Bibliothek durchgeführt. Dabei traten erhebliche Probleme auf, da die gängigen Erschließungsstandards für digitale Medien gezielte Abfragen mit Blick auf die Bedürfnisse der Langzeitarchivierung nur unzureichend erlauben. Vor diesem Hintergrund wurden die internationalen Entwicklungen auf dem Gebiet der Metadaten für die LZA in die Diskussion um die Erschließung digitaler Ressourcen eingebracht. Bei der technischen und inhaltlichen Analyse des Bestandes wurde darüber hinaus festgestellt, dass für die Archivierung datenträgergebundener Medien klare Selektionskriterien definiert werden müssen, da nicht alle E-Ressourcen archivierungswürdig sind. Ein solcher Kriterienkatalog hat auch in die hauseigenen Sammelrichtlinien für digitale Medien Eingang gefunden.<sup>17</sup> Dem Institut für Softwaretechnologie mit dem Team um Prof. Dr. Borghoff ist es im Rahmen dieses Projektes gelungen,

mit Hilfe einer Sammlung alter Hard- und Software repräsentative Dokumente von ihren zum Teil veralteten Datenträgern (8-Zoll-Floppys, 5,25- und 3,5-Zoll-Disketten, CDs, DVDs) zu lösen und in einer modernen Datenbank zu verwalten. Damit findet eine technische und altersmäßige Normierung statt. Ferner bringt der Einsatz eines Standarddatenbanksystems eine vielseitige und vereinheitlichende Unterstützung bei den Managementaufgaben.<sup>18</sup>

Die Erfahrungen aus diesem Projekt wollen beide Partner gemeinsam wieder aufgreifen, um Bestände der Bayerischen Staatsbibliothek von den zum Teil veralteten Datenträgern zu lösen und in einer Standarddatenbank zu verwalten. Die geeignete Auswahl der Bestände vor dem Hintergrund des Erwerbungs- und Sammelprofils sowie rechtlicher Aspekte, die Erweite-

**Klärung rechtlicher Fragen  
mit den Produzenten**

**Projektschwerpunkt:  
Workflows**

rung der Erschließungsparameter sowie die notwendigen Verfahren zum dauerhaften Erhalt der Verfügbarkeit der Informationen sind Gegenstand der Untersuchung. Dabei ist geplant, für alle Arbeitsprozesse von der Auswahl, der Loslösung, der Erschließung bis hin zur Bereitstellung eine Aufwandsabschätzung durchzuführen. Aus den Erfahrungen der retrospektiven Bearbeitung sollen Schlüsse für die Behandlung des laufenden Zugangs an digitalen Medien abgeleitet werden.

## SCHLUSSBEMERKUNG

Die Bayerische Staatsbibliothek steht, eingebunden in überregionale und nationale Initiativen, vor vielfältigen Herausforderungen auf dem Feld der Langzeitarchivierung elektronischer Medien. Aufgrund ihrer Aufgabenvielfalt auf den verschiedensten Ebenen sind zahlreiche rechtliche und technische Probleme, aber auch Fragestellungen bezüglich der Auswahlprofile und der Nutzungsmodelle zu lösen. Sie beteiligt sich aktiv an der Entwicklung kooperativer Lösungen für Deutschland und leistet ihren Beitrag bei der konkreten Archivierungspraxis. Mit der Entwicklung eines Workflows für den Bereich der Retrodigitalisate ist ein Anfang auf diesem Feld gemacht. Nun geht es zunächst im Rahmen des skizzierten DFG-Projekts darum, auch für andere Medientypen und Gattungen zu nachhaltigen Konzepten zu gelangen.

<sup>1</sup> Vgl. dazu auch [www.bsb-muenchen.de/ejournal/konsort.htm](http://www.bsb-muenchen.de/ejournal/konsort.htm).

<sup>2</sup> Vgl. dazu die in der Bayerischen Landesbibliothek Online (BLO) erschlossenen landesspezifischen Internetquellen ([www.bayerische-landesbibliothek-online.de](http://www.bayerische-landesbibliothek-online.de)).

<sup>3</sup> Dies geschieht derzeit in den Bereichen Geschichte (<http://mdz2.bib-bvb.de/hist/>) und Osteuropa ([www.vifaost.de/sys/cgi/w/index.cgi?p=fachuebergreifend/internetressourcen](http://www.vifaost.de/sys/cgi/w/index.cgi?p=fachuebergreifend/internetressourcen)) im Rahmen des Fachportals Chronicon ([www.chronicon.de/](http://www.chronicon.de/)) bzw. der Virtuellen Fachbibliothek Osteuropa ([www.vifaost.de/](http://www.vifaost.de/)). Für die Fächer Klassische Altertumswissenschaften und Musikwissenschaft befindet sich der Aufbau entsprechender Internetquellensammlungen in Vorbereitung.

<sup>4</sup> Vgl. dazu Berg, Heinz-Peter; Schäffler, Hildegard; Sens, Irina: Elektronische Zeitschriften in der überregionalen Literaturversorgung: Ergebnisse des DFG-Projektes EZUL. In: ZFB 49:3 (2002), S. 118–132.

<sup>5</sup> Weitere Informationen zu dem Projekt »Überregionale Bereitstellung und Förderung von Online-Datenbanken im Bereich geisteswissenschaftlicher Sondersammelgebiete« finden sich unter [www.bsb-muenchen.de/daten/ppuprojekt.htm](http://www.bsb-muenchen.de/daten/ppuprojekt.htm) bzw. in Horstkemper, Gregor: »Nosse volunt omnes, mercedem solvere nemo? Pay-per-Use als Instrument der überregionalen Informationsversorgung. Workshop an der Bayerischen Staatsbibliothek. In: Bibliotheksdienst 38:11 (2004), S. 1457–1469.

<sup>6</sup> Vgl. dazu auch den Beitrag von Marianne Dörr in diesem Heft.

<sup>7</sup> Vgl. dazu Wiesenmüller, Heidrun; Jendral, Lars; Schäffler, Hildegard; Schomburg, Silke; Höck, Christoph; Kohlmann, Monika: Auswahlkriterien für das Sammeln von Netzpublikationen im Rahmen des elektronischen Pflichtexemplars: Empfehlungen der Arbeitsgemeinschaft der Regionalbibliotheken. In: Bibliotheksdienst 11 (2004), S. 1423–1444.

<sup>8</sup> Vgl. dazu den Beitrag von Dobratz, Neuroth, Schoger in diesem Heft.

<sup>9</sup> RLG/OCLC Working Group on Digital Archive Attributes: Trusted Digital Repositories: Attributes and Responsibilities: An RLG-OCLC Report. Mountain View, CA: RLG, 2002. [www.rlg.org/en/pdfs/repositories.pdf](http://www.rlg.org/en/pdfs/repositories.pdf).

<sup>10</sup> Eine ausführliche Beschreibung des Projektes nestor und des Arbeitsschwerpunktes der BSB findet sich auf den Webseiten des Projektes unter [www.langzeitarchivierung.de](http://www.langzeitarchivierung.de).

<sup>11</sup> Die neue Ausstattung in der Digitalisierungsstraße ermöglicht Farbdigitalisierungen von großformatigen Vorlagen (maximale Auflösung von 400 dpi bei Ao bzw. 600 dpi bei A1) und Handschriften, was rasch zu sehr großen Datenmengen führen wird.

<sup>12</sup> Eine Übersicht der Digitalen Sammlungen der BSB findet sich unter <http://mdz.bib-bvb.de>.

<sup>13</sup> Vgl. Deutsche Forschungsgemeinschaft: Bericht der Arbeitsgruppe Technik zur Vorbereitung des Programms »Retrospektive Digitalisierung von Bibliotheksbeständen« im Förderbereich »Verteilte Digitale Forschungsbibliothek«. [www.sub.uni-goettingen.de/ebene\\_2/vdf/endfass1.htm](http://www.sub.uni-goettingen.de/ebene_2/vdf/endfass1.htm); Deutsche Forschungsgemeinschaft: Die Erschließung und Bereitstellung digitalisierter Drucke: Vorschläge des Unterausschusses für Kulturelle Überlieferung. [www.dfg.de/forschungsfoerderung/wissenschaftliche\\_infrastruktur/lis/download/konzept\\_digitale\\_drucke.pdf](http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/konzept_digitale_drucke.pdf); Deutsche Forschungsgemeinschaft: Praxisregeln im Förderprogramm »Kulturelle Überlieferung«. [www.dfg.de/forschungsfoerderung/formulare/download/12\\_151.pdf](http://www.dfg.de/forschungsfoerderung/formulare/download/12_151.pdf). Siehe auch: Boudrez, Filip; Dekeyser, Hannelore: DAVID – Digitaal archiveren in de praktijk: Handboek. [www.antwerpen.be/david/index.htm](http://www.antwerpen.be/david/index.htm); Dörr, Marianne: Retrodigitalisierung in Bibliotheken und wissenschaftlichen Einrichtungen. <http://computerphilologie.uni-muenchen.de/jgo1/doerr.html>.

<sup>14</sup> Schmidt, Wilhelm R.: Bestanderhaltung als Garant des kulturellen Gedächtnisses in Deutschland. [www.uni-muenster.de/Forum-Bestanderhaltung/grundlagen/allg-schmidt.shtml](http://www.uni-muenster.de/Forum-Bestanderhaltung/grundlagen/allg-schmidt.shtml).

<sup>15</sup> Zur Text Encoding Initiative vgl. [www.tei-c.org/](http://www.tei-c.org/).

<sup>16</sup> Vgl. dazu den Beitrag von Kathrin Schroeder in diesem Heft.

<sup>17</sup> Weitere Informationen zum Projekt »Langzeitarchivierung digitaler Medien« an der BSB finden sich unter [www.bsb-muenchen.de/mdz/izarch.htm](http://www.bsb-muenchen.de/mdz/izarch.htm).

<sup>18</sup> Ausführliche Informationen zum Projekt »Langzeitarchivierung digitaler Medien« an der Universität der Bundeswehr finden sich unter <http://forschung.unibw-muenchen.de/ainfo.php?id=4177>.

Alle URLs Stand 28.04.2005.

## DIE VERFASSER

**Dr. Hildegard Schäffler** ist Leiterin des Referats Zeitschriften und Elektronische Medien der Bayerischen Staatsbibliothek, 80328 München, E-Mail: schaeffler@bsb-muenchen.de

**Dr. Astrid Schoger** ist Leiterin des Sachgebiets Langzeitarchivierung im Referat Digitale Bibliothek der Bayerischen Staatsbibliothek, 80328 München, E-Mail: schoger@bsb-muenchen.de

**Dr. Margarete Wittke** ist verantwortlich für den Bereich Massendigitalisierung im Referat Digitale Bibliothek der Bayerischen Staatsbibliothek, 80328 München, E-Mail: wittke@bsb-muenchen.de