

Die Migration der VD-17-Datenbank vom System OMNIS-Myriad zu PICA

The »Index of Printed Material Published in German-Speaking Countries in the 17th Century (VD 17)« was created and maintained with the database system OMNIS-Myriad until September 2002. This report explains why a change in software was necessary after seven years and elaborates in detail on the decision-making process, the preparatory phase and the steps involved in the migration to the PICA system. Particular emphasis is given to those aspects of the project which involved changes in category format and dataset structures, as well as to the cataloguing process. In conclusion, it is reported that the experience with the migration of the VD-17 database was very positive.

Das »Verzeichnis der im deutschen Sprachraum erschienenen Drucke des 17. Jahrhunderts (VD 17)« wurde bis September 2002 im Datenbanksystem OMNIS-Myriad erstellt. Im vorliegenden Aufsatz wird dargelegt, weshalb nach sieben Jahren ein Systemwechsel notwendig war. Die Entscheidungs-, Vorbereitungs- und Arbeitsphasen bei der Durchführung der Migration zum PICA-System werden detailliert beschrieben. Ein Schwerpunkt liegt auf der Darstellung der Aspekte, die die Änderungen im Kategorienformat und den Satzstrukturen und die Arbeitsgänge bei der Erschließung betreffen. Als Fazit wird festgehalten, dass die Migration der VD-17-Datenbank insgesamt sehr positiv verlaufen ist.

EINLEITUNG

Das »Verzeichnis der im deutschen Sprachraum erschienenen Drucke des 17. Jahrhunderts (VD 17)« ist ein nationalbibliographisches Gemeinschaftsprojekt von mittlerweile neun Bibliotheken unter Federführung der Bayerischen Staatsbibliothek. Ziel des Projektes ist, die Lücke in der bibliographischen Verzeichnung der Druckproduktion im deutschen Raum für diesen Zeitraum zu schließen. Berücksichtigt werden alle deutschsprachigen Titel und, unabhängig von ihrer Sprache, alle im historischen deutschen Sprachgebiet gedruckten und verlegten Werke zwischen 1601 und 1700 (mit Ausnahme von Musikdrucken und Karten).

Die Deutsche Forschungsgemeinschaft (DFG) hatte Anfang der 1990er Jahre einen kooperativen Antrag der Bayerischen Staatsbibliothek München, der Herzog August Bibliothek Wolfenbüttel, der Staatsbibliothek zu Berlin und des Forschungszentrums für Wissensbasierte Systeme (FORWISS) in München angeregt. Nach dem erfolgreichen Abschluss der knapp zwei Jahre dauernden Vorarbeiten kamen aber bereits zu Beginn der Hauptphase am 1.7.1996 die Universitäts- und Landesbibliothek Sachsen-Anhalt Halle, die Universitäts- und Forschungsbibliothek Erfurt/Gotha und die Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden als weitere Bibliotheken hinzu. In den Jahren 1998/1999 konnten außerdem die Daten der Ratsschulbibliothek Zwickau eingespielt

werden. Seit Dezember 2002 gehören die Herzogin Anna Amalia Bibliothek Weimar, seit Anfang 2003 die Niedersächsische Staats- und Universitätsbibliothek Göttingen und seit Mitte 2003 die Stadtbibliothek Nürnberg zum Kreis der teilnehmenden Bibliotheken. Die Hauptphase des Projekts, für die die DFG die Initialförderung übernimmt, ist auf zehn bis zwölf Jahre, beginnend im Jahre 1996, konzipiert.

Bis Ende 2003 konnten bereits circa 190.000 Titel (von den insgesamt geschätzten bzw. hochgerechneten etwa 300.000 Titeln) katalogisiert werden. Die Datenbank findet seit der Öffnung für die Öffentlichkeit im Jahr 1998 bei den wissenschaftlichen Nutzern im In- und Ausland als wichtige bibliographische Informationsquelle breite Akzeptanz.

DIE VD-17-DATENBANK VOR DER MIGRATION UNTER DEM SYSTEM OMNIS/MYRIAD

Die wichtigsten Systemanforderungen an eine VD-17-Datenbank sind: Kooperation von Partnern aus verschiedenen Bibliotheksverbänden, Verknüpfung der Titelaufnahmen mit den digitalen Bildinformationen von ausgewählten Schlüsselseiten und allgemeine Verfügbarkeit der Datenbank am Arbeitsplatz des Nutzers mit umfassenden und komfortablen Recherchemöglichkeiten. In der Konzeptions- und Vorbereitungsphase des Projekts stand kein adäquates Bibliothekssystem zur Verfügung, das alle Anforderungen erfüllt hätte. Deshalb wurde eine damals innovative Softwarelösung gewählt, die aus zwei Teilen bestand: Allegro (das in der Herzog August Bibliothek für die Katalogisierung innerhalb der Sammlung Deutscher Drucke im Einsatz war) als Erfassungskomponente und OMNIS/Myriad für die Text- und Bilddatenverwaltung. Die Überlegung, eine eigene Erfassungskomponente für OMNIS/Myriad zu entwickeln, war aufgrund des zu erwartenden Aufwands zugunsten der Portierung des Allegro-Systems von DOS zu UNIX verworfen worden.² Während das Vorläuferprojekt VD 16³ noch ohne Hilfe der elektronischen Datenverarbeitung durchgeführt wurde,⁴ stellte die gewählte Lösung für das VD 17 einen sehr großen Fortschritt dar. Die Speicherung von digitalisierten Bildern als zusätzliche Erschließungselemente war zur damaligen Zeit ein Novum, und das VD 17 stand diesbezüglich als erstes größeres Projekt an der Spitze der technologischen Entwicklung.



Werner Holbach

zweiteilige Softwarelösung mit Allegro und OMNIS/Myriad

OMNIS/Myriad baute auf einem kommerziellen relationalen Datenbanksystem der Firma TransAction auf, das für den VD-17-Einsatz aber eine anwendungsspezifische Weiterentwicklung durch das Forschungszentrum für Wissensbasierte Systeme (FORWISS) erfuhr. So war unter anderem die Schnittstelle zu dem vorhandenen Allegro-Erfassungssystem zu entwickeln. Zusätzlich musste eine zweite relationale Datenbank konzipiert werden, in der die hierarchische Struktur der MAB-Daten⁵ gespeichert war. Dies war notwendig, um einen Export im MAB-Format für eine spätere Belieferung der Verbunddatenbanken zu ermöglichen. Aus ihr wurden die Informationen zudem in ein flaches OMNIS-Dokument überführt, wodurch die besonderen Volltextretrievalmöglichkeiten des Systems zum Tragen kamen. Allegro und das OMNIS-System wurden so aneinander gekoppelt, dass Datenexport und -import für die Katalogisierer einfach möglich war und alle Sicherungskonzepte des darunter liegenden Datenbanksystems voll ausgenutzt werden konnten. Da mehrere Bibliotheken über das Internet zusammenarbeiteten, musste außerdem ein Synchronisationskonzept entwickelt werden, das Mehrfachfassungen verhinderte. Die digitalisierten Bilder wurden mit ihren Attributen, wie Identifikatoren, Auflösung, Größe, Kompressionsverfahren etc., in mehreren Bilddatenbanken gespeichert. Die Textdatenbank und die angegliederten Bilddatenbanken stellten jeweils eigenständige Komponenten dar. So konnte das Bildmaterial getrennt vom täglichen Katalogisierungsbetrieb erstellt und verwaltet werden. Die dazu notwendige Software war eine Neuentwicklung von FORWISS.⁶ Neben dem OMNIS-Client wurde auch ein WWW-Gateway zur Recherche geschaffen.

Der Arbeitsablauf unter der damaligen Systemkonfiguration soll im Folgenden kurz geschildert werden, weil dadurch auch die Vorteile und die Defizite aus heutiger Sicht deutlich werden: Die Projektbearbeiter recherchierten zunächst in der VD-17-Datenbank, ob zu dem vorliegenden Buch bereits eine Titelaufnahme in der Datenbank vorhanden war. Wenn ja, wurde diese Aufnahme in den Erfassungs-Client geladen, gegebenenfalls erweitert oder modifiziert und ein neuer Lokaldatensatz angehängt. Während der Bearbeitung war der entsprechende Satz für den Zugriff durch andere Bibliotheken gesperrt. Unmittelbar nach Beendigung der Arbeit und Rückimport in die zentrale Datenbank stand der Datensatz wieder zur Bearbeitung durch andere zur Verfügung.

Es gab keine gestufte Rechtevergabe für die Bearbeitung; alle Katalogisierer waren gleichberechtigt. Prinzipiell konnte jede Bibliothek alles in einer Aufnahme, die von einer anderen Bibliothek in die Datenbank

eingebraucht wurde, ändern. In der Praxis sah das so aus, dass bei Korrekturvorhaben mit der Bibliothek, die die Titelaufnahme in die Datenbank eingebracht hatte, Rücksprache gehalten wurde. Nur triviale Fehler (Tippfehler und Ähnliches) wurden ohne Nachfrage bei der betroffenen Bibliothek verbessert.

War noch keine Aufnahme vorhanden, recherchierte der Bearbeiter in einem separaten, aber mit der gleichen Datenbanksoftware verwalteten Fremddatenpool, der die für den Erscheinungszeitraum 17. Jahrhundert relevanten Daten des DBI-VK⁷ enthielt. Dort gefundene Aufnahmen, die hauptsächlich aus den Projekten der Altbestandskonvertierung stammten, konnten genauso wie bereits in der Datenbank vorhandene VD-17-Katalogisate als Vorleistung für die Katalogisierung übernommen werden.

Die Übernahme neu erstellter Katalogisate in die Datenbank lief ohne Dublettencheck. Die Erfassung erfolgte inklusive der Eingabe der vorangestellten Kategoriennummern in einem einfachen Editorfenster. Bei der Eingabe fanden so gut wie keine Plausibilitätsprüfungen statt, so dass es möglich war, wichtige Kategorien zu vergessen bzw. doppelt zu vergeben.

Die alte Datenbank bot keine Verknüpfungen zu Normdateien. Bei der Katalogisierung mussten die Personennamen daher in der externen überregionalen Personennamendatei (PND) recherchiert und gegebenenfalls die entsprechenden Namen per Cut-and-Paste-Verfahren in die Erfassungsmaske übertragen werden. Die PND-Nummern wurden nicht in die VD-17-Datenbank übernommen.⁸ Noch nicht in der PND enthaltene Namen wurden recherchiert und offline an die zuständige PND-Redaktion in der Bayerischen Staatsbibliothek gemeldet. PND-Korrekturen bzw. Ansetzungsänderungen mussten im VD 17 manuell an allen betroffenen Datensätzen nachvollzogen werden.

Vor der Erfassung von Druckern/Verlegern bzw. Druckorten war in einer separaten Allegro-Datenbank der Drucker und Verleger zu recherchieren. Fand sich dort eine entsprechende Ansetzungsform, wurde sie ebenfalls per Cut-and-Paste übernommen. Neue Namen mussten offline gemeldet werden.

Neben den Textdaten gibt es im VD 17 Bilddaten zu bestimmten Schlüsselseiten der Drucke. Die durchschnittliche Zahl liegt bei etwa drei Images; in Einzelfällen können einem Katalogisat aber auch 20 oder mehr Bildseiten zugeordnet sein. Die Digitalisierung erfolgte nach der Katalogisierung. Als Dateiname wurde die Datensatz-Produktionsnummer⁹ (auch ID-Nummer genannt) mit einigen weiteren Informationen (Kennzeichnung von Bildtyp, laufende Nummer, formatspezifische Endung) verwendet. Die Bilder wurden systembedingt in mehreren, separaten Bilddaten-

banken verwaltet. Die Verknüpfung zwischen Text- und Bilddatenbanken erfolgte über ein Programm, das die Bilddatenbanken an der Textdatenbank »anmeldete«. Die Links zu den Bildern in der Titelanzeige wurden dynamisch generiert, d. h., dass die Katalogisate keine URLs enthielten und keine manuellen Eingaben bzw. Änderungen von URLs notwendig waren.

Für Text- und Bilddatenbanken wurden jeweils eine interne Arbeits- und eine externe öffentliche Version vorgehalten. Die Textdatenbanken waren von der Struktur her identisch. Bei den Bilddatenbanken unterschieden sich die Versionen in den angebotenen Formaten (öffentlich GIF und JPEG mit 150 dpi, intern TIFF und JPEG mit 300 dpi).

Die OMNIS-Datenbank bot neben der strukturierten (feldbezogenen) Suche die Möglichkeit der Volltextsuche in prinzipiell allen Elementen der Titelaufnahme (inklusive Kollation, Fußnoten etc.). Als weitere Recherchemöglichkeiten standen die Stringsuche und die Möglichkeit der Verknüpfung mit Booleschen Operatoren sowie die Rechtstrunkierung zur Verfügung. Auch Informationen, die unterschiedlichen Satztypen (h-Satz mit übergeordneten Titelinformationen, u-Satz mit bandbezogenen Informationen, l-Satz mit exemplarbezogenen Informationen) entstammten, waren frei kombiniert suchbar. Die Kombination von Strukturfeld- und Volltextsuche war möglich.

Diese Suchmöglichkeiten in der VD-17-Datenbank beruhten auf einer redundanten Datenhaltung. Jedes Katalogisat wurde als Volltextinformation abgespeichert (darauf basierte die Volltextsuche) und als Strukturinformation. Als drittes Element wurden die Katalogisate präformatiert als statische HTML-Seite in der Datenbank abgelegt. Dies führte einerseits zu einer schnelleren Anzeige der vollen Information, andererseits kam es beim Import von neuen Katalogisaten, aber auch nach Korrekturen oder Ergänzungen, zu einer gewissen Verzögerung, bis der so genannte »Update-Manager« aus den importierten Daten das HTML-Dokument aufgebaut hatte.

Bei der Anzeige von Suchergebnissen im Online-Katalog wurde für die Vollanzeige die ID- bzw. Produktionsnummer des Satzes und die Zahl der verfügbaren Bilder aus der Datenbank gewonnen und der Rest der Titelaufnahme aus dem statischen HTML-Dokument¹⁰ genommen. Es war aber auch möglich, die Strukturfelder eines Katalogisats anzusehen. Diese Anzeige wurde dynamisch aus der Datenbank erzeugt.

Die Anzahl der zu einem Dokument vorhandenen Bilder wurde in dem oberen, dynamisch generierten Teil der Dokumentenanzeige durch fortlaufende Nummern angezeigt. Durch Anklicken der Hyperlinks öffneten sich die Bilder der öffentlichen Datenbank im

Browser; für die Arbeitsdatenbank wurde ein Viewer gestartet, da hier vorwiegend TIFF-Bilder vorlagen, die von bisherigen Browser-Generationen nicht unterstützt werden.

Die öffentliche Datenbank konnte bereits in die Portale »Elektra« der Bayerischen Staatsbibliothek und »Gateway Bayern« des Bibliotheksverbundes Bayern eingebunden werden.

ENTSCHEIDUNG FÜR DIE MIGRATION UND EIN NEUES SYSTEM

OMNIS/Myriad stellte anfangs ein sehr fortschrittliches System dar; es bewährte sich nach Projektbeginn für längere Zeit. Im Laufe der Jahre zeigten sich aber auch zunehmend Schwachstellen, die eine kontinuierliche Weiterentwicklung erfordert hätten. Performance-Probleme, insbesondere beim Im- und Export der Datensätze, nahmen zeitweise stark zu.¹¹ Eine Ausweitung des Teilnehmerkreises wäre ohne grundlegende Neuprogrammierung nicht möglich gewesen. FORWISS war jedoch als universitätsnahe Forschungs- und Entwicklungsinstitution nicht darauf ausgelegt, eine Dauerpflege und systematische kundengerechte Weiterentwicklung zu leisten. Die notwendige Zukunftssicherheit für den produktiven Betrieb des VD-17-Projekts musste aber erreicht werden. Ein Einfrieren des Systems ohne Weiterentwicklung war aufgrund des schnellen technologischen Wandels nur für eine begrenzte Zeit tragbar.

Auch die Tatsache, dass mit einem eigenen Erfassungssystem (Allegro-X) gearbeitet wurde, hatte immer wieder mehr oder minder große Probleme bereitet. Dringend notwendige Weiterentwicklungen, wie z. B. die Einbindung einer verknüpften Personennamendatei und Drucker-/Verlegerdatei, schienen auf dieser Plattform nicht mehr sinnvoll realisierbar. Auch die Bilddatenverwaltung hatte, bedingt durch viele manuelle oder halbautomatische Arbeitsschritte und fehlende Integritätskontrollen, Schwierigkeiten bereitet. Des Weiteren erforderte die WWW-Datenbank eine regelmäßige Offline-Versorgung; eine Öffnung der Arbeitsdatenbank war nicht möglich, da sonst eine Zunahme der Performance-Probleme zu befürchten war. Mit Auslaufen der DFG-Förderung für FORWISS Mitte 2001 endete daher die erfolgreiche Partnerschaft im Rahmen des VD-17-Projekts mit dem Münchner Institut.

Bereits im Jahre 2000 hatte es erste Überlegungen für einen Systemwechsel gegeben. In einer Sitzung der Projektbibliotheken im Februar 2001 wurden die Migrationsüberlegungen konkretisiert. Die Bibliotheken waren sich einig, dass das VD 17 eine eigenständige Datenbank bleiben sollte, da nur dadurch spezielle Konfigurationen einer Fachdatenbank realisierbar

Verknüpfung von Text- und Bilddatenbank

Zukunftssicherheit des bisherigen Systems nicht gewährleistet

erste Migrationsüberlegungen 2000

waren oder auch die Indizierung von normalerweise nicht suchbaren Feldern ermöglicht werden konnte. Die Perspektive einer unproblematischen Datenlieferung an die Verbünde sollte jedoch im Blick behalten werden. Für den Umstieg kamen nur gängige, kommerzielle und dauerhaft gepflegte Standardsysteme in Frage. Die Option einer Weiterpflege der bestehenden VD-17-Software durch eine Firma wurde zwar diskutiert, aber Zukunftssicherheit auch in der Betreuung schien nur ein Bibliothekssystem zu bieten, das bereits in einem Bibliotheksverbund im Einsatz war.

Die Bayerische Staatsbibliothek übernahm die Aufgabe, einen Anforderungskatalog vorzubereiten. Dieser sollte unter den beteiligten Bibliotheken abgestimmt werden und anschließend die Grundlage für die Einholung eines Angebots bei den in Frage kommenden Anbietern sein. Zielzeitpunkt für einen Umstieg war die zweite Jahreshälfte 2002.

Im Frühjahr 2001 wurde daraufhin der Leistungskatalog für das neue VD-17-System erstellt. Die zentralen Anforderungen waren:

**zentrale Anforderungen
an das neue System**

Umfassende, komfortable Katalogisierungsmöglichkeiten unter Berücksichtigung des MAB2-Formates und Verwendung von Dublettencheck und Plausibilitätskontrollen
Integration von Images
Möglichkeit der Nutzung von Fremddaten
Differenzierte Rechtevergabe
Verknüpfung zu Normdateien
Volltextsuche und strukturierte Suche
Benutzerfreundliche, anpassbare Web-Oberfläche mit Imageanzeige

Nach Abstimmung des Anforderungskatalogs wurden in der zweiten Hälfte des Jahres 2001 die Firmen OCLC/PICA (PICA CBS4-System) und ExLibris (Aleph- und DigiTool-System) um einen Umsetzungsvorschlag und Testplan sowie um ein Angebot gebeten.

Das Ergebnis war, dass beide Firmen fachlich überzeugende Angebote vorlegten, die jedoch in den Preisvorstellungen differierten. Dies lag nicht so sehr an den Preisen der Server-Lizenz, die im Volumen vergleichbar waren, sondern vor allem im Bereich der Client-Lizenzen. Hier wären bei einer Entscheidung für ExLibris zusätzliche und recht erhebliche Kosten pro Arbeitsstation für die Bibliotheken hinzugekommen. Dies hätte für jede Einzelbibliothek und natürlich für das Gesamtprojekt eine ganz erhebliche Belastung dargestellt. Bei PICA hingegen fehlte dieser Kostenfaktor, da alle Bibliotheken über die Mitarbeit bei ZDB und PND bereits zur Nutzung der Zugriffssoftware WIN-IBW berechtigt waren. Außerdem wäre Ex-

Libris zum damaligen Zeitpunkt eine Insellösung gewesen, die beträchtliche Personalkosten für Betreuung und Betrieb erfordert hätte. Aus diesen Gründen kristallisierte sich eine erste Präferenz für PICA heraus.

Im Januar 2002 schlug die Bayerische Staatsbibliothek vor, dass eine Teststellung eines PICA-Systems mit echten VD-17-Daten erfolgen und bei zufriedenstellendem Verlauf eine PICA-Lösung beauftragt werden sollte. Die Bibliotheken sprachen sich einhellig für dieses Vorgehen aus. Für die nicht im Gemeinsamen Bibliotheksverbund in Göttingen (GBV) vertretenen Bibliotheken dürfte aber kein Nachteil entstehen und eine langfristig tragfähige Administrationsbasis für die Datenbank müsste gefunden werden. Aus diesem Grund wurde angeregt, eine Installation beim GBV und nicht bei PICA in Leiden (Niederlande) oder an der Bayerischen Staatsbibliothek ins Auge zu fassen. Die Bayerische Staatsbibliothek als bisherige Betreiberin der VD-17-Datenbank erklärte sich einverstanden. Die Tests beim GBV verliefen erfolgreich und so wurde auf der Sitzung der VD-17-Bibliotheken am 2. Mai 2002 in Göttingen die Migration des VD 17 auf eine PICA-Plattform einstimmig beschlossen. Die Verbundzentrale des GBV wurde gebeten, die Migration durchzuführen und das Hosting der neuen VD-17-Datenbank zu übernehmen.

Die Verbundzentrale legte im Mai 2002 ein Angebot für den Betrieb der Datenbank vor. Dieses umfasste die Punkte »Daten- und Verfahrensmigration« und »Hardware für Imagedatenbank« mit einmaligen sowie »OCLC/PICA-Lizenzen« und »Betriebskosten Verbundzentrale« als laufende jährliche Kosten. Die Hardwarekosten enthielten den Aufwand für etwa vier GB Plattenspeicher (RAID-Array), was dem geschätzten Bedarf für die nächsten vier Jahre bei Doppelspeicherung von TIFF- und GIF-Bildern entsprach, sowie die Beschaffung eines adäquaten Servers. Die Titeldatenbank erforderte keinen zusätzlichen Hardwareinsatz, da sie auf vorhandenen Servern des GBV läuft. Die Kosten für die Daten- und Verfahrensmigration (inklusive Erst-Einrichtung der Software- und Serverbetriebsumgebung) berechneten sich aus der aufzuwendenden Arbeitszeit. Die OCLC/PICA-Lizenz umfasste die Nutzung des Katalogisierungssystems und der WWW-Datenbank ohne Zugangslimitierung. Die Betriebskosten schließlich wurden auf der Basis eines geschätzten Aufwandes berechnet. Diese schließen Softwarepflege, Serverbetriebsystem, Serverbetrieb, Entwicklung und Betrieb Datendienst, Datenkorrekturen kleineren Umfangs, Images-Verwaltung etc. ein. Das von der Göttinger Verbundzentrale vorgelegte Angebot wurde angenommen.

Auf der Göttinger Sitzung war weiterhin beschlos-

Präferenz für PICA

sen worden, eine Arbeitsgruppe aus Vertretern der Katalogisierungsarbeitsgruppen aller VD-17-Bibliotheken und der Verbundzentrale des GBV einzurichten. Aufgabe der Gruppe war es, ein Anforderungspapier zu erarbeiten. Außerdem wurde der weitere Zeitplan für den Migrationsverlauf (mit Zielvorstellung Ende September 2002 für den Umstieg) festgelegt.

Die Domain »vd17.de« sollte weiter in München gepflegt werden. Ein Link auf die neue Göttinger Datenbank-Adresse sollte zur Recherche in der Web-Datenbank führen.

Die Bayerische Staatsbibliothek stellte einen Förderantrag auf Übernahme der Kosten an die DFG, der bewilligt wurde. Er umfasste die Finanzierung der Kosten für die Daten- und Verfahrensmigration, die Kosten für die Hardware-Neubeschaffung und den Ersterwerb einer OCLC/PICA-Lizenz. Als Eigenleistung wurde von den Bibliotheken zugesagt: Neuausstattung der Arbeitsplätze mit Standard-PCs, Installation und Betreuung der Arbeitsplatz-Software¹² durch die lokalen IT-Abteilungen und die Übernahme der Betriebskosten der Verbundzentrale und der jährlichen Lizenzkosten für PICA. Wegfallen konnte durch erheblich reduzierten personellen Aufwand außerdem die bis dahin durch die DFG finanzierte Systembetreuer-Stelle. Die Gründe sind der Einsatz einer Standard-Verbundsystem-Software, die zwar an Projektspezifika angepasst wurde, aber nicht mehr speziell für das Projekt entwickelt wird, sowie der Host-Betrieb bei einer Verbundzentrale, die ohnehin entsprechende Betreuungskapazität vorhalten muss.

Die VD-17-Bibliotheken München, Wolfenbüttel und Berlin sagten als Erstantragsteller und Hauptträger des Projekts zu, diese laufenden Kosten gemeinschaftlich zu tragen und damit auch über die Projektphase hinaus den Bestand und die Verfügbarkeit des VD 17 zu gewährleisten.

ABLAUF DER MIGRATION NACH DER ENTSCHEIDUNG FÜR PICA

Mit dem Beschluss der VD-17-Bibliotheken vom Mai 2002 konnte die Migration in die entscheidende Phase gehen. Zunächst wurde ein detaillierter Projektplan erstellt, der in den nächsten Monaten Schritt für Schritt umgesetzt werden sollte. Als erstes konstituierte sich die Arbeitsgruppe Migration unter Leitung der Herzog August Bibliothek Wolfenbüttel. Die Vertreter aus den beteiligten Bibliotheken und der Verbundzentrale des GBV erarbeiteten in einer gemeinsamen Sitzung Ende Mai in Wolfenbüttel und anschließend per Mailingliste ein Anforderungspapier.

Es wurden im Einzelnen folgende Punkte geklärt bzw. detailliert:

Format-Konkordanz für die Umsetzung

Ein erster, von der Verbundzentrale erstellter Entwurf für eine Konkordanz der Felder im Titeldaten- und Lokaldatenbereich zwischen OMNIS- (bzw. MAB-) Kategorien und PICA-Kategorien lag bereits vor. Ein großer Teil der Vorschläge wurde von der Arbeitsgruppe bestätigt; bei einigen Kategorien wurde Änderungsbedarf gesehen. Ein Punkt betraf die VD-17-Nummern: Da sie wichtig für das Zitieren der Titel sind, sollten die bisherigen VD-17-Nummern beibehalten werden. Für die in MAB-Feld 035 gelieferten Nummern wurde ein neues Feld in PICA geschaffen (PICA 2191).¹³ Um die Struktur der VD-17-Nummern auch zukünftig beibehalten zu können, wurde die Einrichtung eines VD-17-Nummern-Generators beschlossen: Er gewährleistet, dass die Syntax gleich bleibt.¹⁴

Ein weiterer wichtiger Punkt war das Feld PICA 4000, in dem Hauptsachtitel, Zusatz zum Sachtitel, Verfasserangabe und gegebenenfalls Parallelsachtitel enthalten sind. Hier musste geklärt werden, ob die Länge des Feldes für die eher sehr langen Titelfassungen des 17. Jahrhunderts ausreichend ist. Es zeigte sich, dass dies (bis auf ganz wenige Titel) der Fall war.

Auch einige neue Kategorien wurden vereinbart. Körperschaftliche Widmungsempfänger wurden beispielsweise in der alten Datenbank bei den persönlichen Widmungsempfängern und der alternative Fingerprint in den Fußnoten erfasst. In der neuen Datenbank gibt es dafür nun separate Felder.

Satzstrukturen und Modifikationen der satzbezogenen Datenhaltung

Die übergreifende Suche über Gesamtaufnahme und Bandsätze bei mehrbändigen Werken ist nur durch Doppelung in der neuen Datenbank möglich. Auch einige Felder für exemplarbezogene Daten müssen in den Titelsatz kopiert, also gedoppelt werden, da nicht über Titel- und Exemplarsatz gleichzeitig gesucht werden kann. Das betrifft u. a. das Feld für die Provenienz: der Ex-Libris-Vermerk ist nun beispielsweise zusätzlich im Titelsatz in einer Fußnote unterzubringen.¹⁵

Ähnlich verhält es sich mit den Hinweisen zur digitalisierenden Bibliothek. Dieser Eintrag erfolgt zukünftig im Titelsatz, da er prinzipiell von jeder Bibliothek bearbeitet werden können muss. Die dazu nötigen Rechte sind im PICA-System aber nur für die bibliographischen Daten gegeben.

Die Arbeitsgruppe diskutierte auch die Behandlung der Nachsätze (für beigefügte und enthaltene Werke) im PICA-System. Die einfache Möglichkeit, dem Verfasser eine Nebeneintragung in den PICA-Feldern 301x zu geben, wurde verworfen, da die Verfasser-Sachtitel-Zuordnung dadurch aufgelöst worden wäre. Die Grup-

Förderantrag an die DFG

Konstituierung der Arbeitsgruppe Migration

pe sprach sich vielmehr für die Einrichtung neuer Felder analog zu der Struktur der MAB2-Felder 800–829 (Nichtstandardmäßige Nebeneintragungen) aus, da diese den Anforderungen an die Recherchemöglichkeiten besser entsprachen. Die Verbundzentrale und einige Bibliotheken präferierten die im GBV übliche Lösung mit eigenständigen Nachsätzen (j-Stufe), um eine leichtere Nachnutzung der Datensätze zu ermöglichen. Eine vorläufige Entscheidung zugunsten der MAB2-konformen Lösung mit »flacher Struktur« fiel im August 2002. Für die im GBV arbeitenden VD-17-Bibliotheken stellte die Verbundzentrale eine Kopierfunktion für die Übernahme von Titeldaten aus dem VD 17 zur Verfügung, die die Daten soweit möglich in die Verbundstruktur umsetzt. Eine endgültige Entscheidung wird erst getroffen, wenn die Richtlinien des GBV für unselbstständig erschienene Werke vorliegen.

Indexierung und Suche

Für die Suche in der VD-17-Datenbank legte die Arbeitsgruppe die Zuordnung der Datenbankkategorien zu den so genannten Suchschlüsseln fest. Es handelt sich dabei um aus drei Großbuchstaben bestehende mnemotechnische Abkürzungen, die bei der Suche mit dem WIN-IBW-Client eingegeben werden müssen bzw. die in der WWW-Datenbank dem Benutzer zur Auswahl angeboten werden. Für den ALL-Index als besonders wichtigen Suchschlüssel war z. B. festzulegen, welche Kategorien für diese Freitextsuche ausgewertet werden sollen. Zusätzlich musste eine Reihe von VD-17-spezifischen Suchschlüsseln vereinbart werden, beispielsweise ZSR für Zensoren, WDM für Widmungsempfänger, GAT für Gattungsbegriffe usw.

Gestaltung der Erfassungsbildschirme und der Anzeige

Die Arbeitsgruppe erstellte Vorgaben zur Gestaltung der Eingabe- und Anzeigebildschirme für den Client und für die Anzeige in der WWW-Datenbank. So wurde festgelegt, welche Kategorien in welcher Reihenfolge und mit welcher Bezeichnung vorkommen sollten. Es stellte sich nach vollzogener Migration heraus, dass hier nachträglich noch einiges verändert werden musste. Dies geschah dann in Absprache mit allen Katalogisierern und der Verbundzentrale. Das Grundprinzip für die Anzeige im WIN-IBW-Client lautete: Es sollen die kompletten Titeldaten plus die Exemplardaten angezeigt werden.

Sonstige Anforderungen

Weiterhin wurden die Fragen der Dubletten- und Plausibilitätskontrollen besprochen. Während Dublettenchecks vorerst nicht vorgesehen sind, wer-

den einige Plausibilitätsprüfungen (z. B. bei Kategoriennummern, Erscheinungsjahr, ...) durchgeführt. Das von der Arbeitsgruppe erstellte Papier wurde in den Herkunftsbibliotheken der Mitglieder geprüft und dann zusammen mit der Verbundzentrale verabschiedet. Die Umsetzung des Anforderungspapiers erfolgte im Juni 2002 zunächst an der bereits vorhandenen kleinen, gut handhabbaren Testdatenbank (inkl. WWW-Zugang) mit VD-17-Daten.

Parallel dazu begann die Verbundzentrale in Zusammenarbeit mit dem VD-17-Systembetreuer mit der Vorklärung technischer Fragen, z. B.:

- Dauer des MAB-Exports der kompletten VD-17-Datenbank für den endgültigen Datenabzug
- Speichervolumen (und Zuwachsraten) für die VD-17-Images
- Modus und Rhythmus der Imagelieferung durch die Bibliotheken und der Einspielung in die Datenbank
- VD-17-Nummern-Generierung inklusive des Algorithmus für Prüfbuchstaben im neuen System.

Weiterhin begann die Arbeitsgruppe, die Katalogisierungsrichtlinien des GBV¹⁶ und die Regeln für die Belegung der Felder für die VD-17-Datenbank in eine neue projektspezifische Arbeitshilfe zu integrieren. Daran waren auch Katalogisierer außerhalb der Arbeitsgruppe beteiligt; die Diskussionen und Abstimmungen hierzu liefen über die Mailingliste. Sie fanden einen (vorläufigen) Abschluss erst einige Zeit nach der Migration.

Die Umsetzung der Daten in der Testdatenbank wurde von den VD-17-Bibliotheken akzeptiert, so dass der Komplettdatenabzug in der zweiten Augushälfte eingespielt werden konnte. In der ersten Septemberhälfte beteiligten sich die Katalogisierer an den Tests. Zu diesem Zeitpunkt waren noch nicht alle Anforderungen berücksichtigt, und es traten weitere Problemfälle auf. Diese wurden im Dialog mit der AG Migration geklärt und die Konvertierungsprogramme von der Verbundzentrale entsprechend modifiziert. Neben den noch nicht umgesetzten Nachsätzen war die Indexierung teilweise falsch bzw. unvollständig. Es fehlten auch die Exemplarsätze und der ALL-Index war noch nicht realisiert.

Der weitere Ablauf war auch bestimmt von dem Ziel, die Migration ohne Supplementlieferungen zu realisieren, d. h. es durfte keine Weitererfassung im alten System während der Datenumsetzung stattfinden. Zum Stichtag 19. September 2002 endete die Katalogisierung in die OMNIS-Datenbank. Vom 20. bis 25. September wurde in München noch einmal ein aktueller Datenabzug erstellt und an den GBV geliefert. Dieser wurde bis Ende September durch die Verbundzentra-

le umgesetzt und anschließend in eine neue Produktionsdatenbank eingespielt. (Die alte Testdatenbank sollte nun als VD-17-Trainingsdatenbank für Schulungen dienen.) Gleichzeitig wurden die Bilddaten transferiert.¹⁷ Parallel dazu fanden vom 23. bis zum 27. September Schulungen zum Arbeiten mit dem WIN-IBW-Client und zum neuen VD-17-Datenbanksystem statt.¹⁸ In den VD-17-Bibliotheken wurden die Arbeitsplätze für den Einsatz des neuen Systems vorbereitet. Dies bedeutete insbesondere einen Wechsel von UNIX-Workstations zu Windows-PCs und die Installation der WIN-IBW (mit den VD-17-eigenen Anpassungen) durch die jeweils lokalen IT-Betreuer.

Nach Abschluss der Schulungen war ab 30. September 2002 die Neukatalogisierung möglich. Allerdings waren anfangs noch nicht alle Altdaten indiziert. Aufwändiger als erwartet gestaltete sich auch die Implementierung eines VD-17-Produktionsnummerngenerators. Zum 1. November 2002 stand schließlich das Katalogisierungssystem mit voller Funktionalität für Neukatalogisierung und für Korrekturen zur Verfügung. Ergänzend zu den Titeldaten wurden im Laufe des Oktobers Druckernormsätze und ein VD-17-relevanter Auszug aus der überregionalen Personennamendatei eingespielt.

NACH DER MIGRATION

Nach Freigabe des Produktionsbetriebs fanden weiterhin umfangreiche Tests statt und es gab eine Reihe von Nachbesserungswünschen von Seiten der Katalogisierer. Diese wurden soweit möglich durch die Göttinger Verbundzentrale maschinell durchgeführt. So zeigten sich auch einige Umsetzungsfehler, von denen die meisten behoben werden konnten und von denen einige beispielhaft aufgeführt sind:

Es gab in einigen wenigen Fällen falsche Zuordnungen von Verweisungsformen auf Namen.

Im Feld Erscheinungsort ging der in eckigen Klammern angegebene Inhalt (z. B. aufgelöste Abkürzungen) verloren.

In der alten Datenbank waren Nebeneintragungen bei Sachtiteln in einem Feld untergebracht. Die Aufspaltung auf mehrere PICA-Felder wurde bei der Migration nicht durchgeführt.

Ebenfalls bei Nebeneintragungen erfolgte die Umsetzung des bisherigen Nichtsortierzeichens – durch das in PICA gültige Nichtsortierzeichen @ nicht.

Bei der Umsetzung von in Vorlageform vorliegenden persönlichen Namen in den Feldern 3060, 3061 und 3062 (Widmungsempfänger, Zensoren, Beiträger) wurde, wenn ein Komma enthalten war, die Namensform als invertierte Form interpretiert.

Einige Probleme bei der Umsetzung gingen auf die

Struktur bzw. Uneinheitlichkeit der Daten zurück, die nur teilweise vor der Migration korrigiert und vereinheitlicht werden konnten.

Einige Änderungs- und Korrekturwünsche konnten erst bei dem Ende Februar 2003 erfolgten Umstieg von der CBS3-Version auf die CBS4-Version umgesetzt werden. So wurde der bis dahin sehr vermisste ALL-Index eingerichtet und die so genannte »su-Anzeige« im Client so gestaltet, dass der Katalogisierer zur Überprüfung der Titelaufnahme sich diese in der Form anzeigen lassen kann, wie sie sich dem Benutzer darstellt.

Parallel zur Katalogisierungsdatenbank wurde ein WWW-Zugang implementiert, den die VD-17-Partner ab Anfang November testeten. Die WWW-Datenbank wird online aktualisiert, so dass die Benutzer ohne Zeitverzug Zugriff auf die aktuellen Daten haben. Nachdem auch das Problem der Skalierung der Bilddateien gelöst werden konnte, erfolgte die Freischaltung der Datenbank für die Öffentlichkeit im Februar 2003. Die Bilder werden nun, sobald sie das erste Mal aufgerufen werden, automatisch auf eine Größe skaliert, die auf dem Bildschirm gut angezeigt werden kann. Dieser Vorgang kann einen kurzen Moment dauern, aber bereits beim zweiten Aufruf erscheint das angepasste Bild sofort. Zusätzlich wurde ein Link eingebaut, der zum vergrößerten Bild in Originalauflösung führt. Mit der WWW-Schnittstelle ist auch die Voraussetzung für eine Einbindung in den Karlsruher Virtuellen Katalog und andere Internetportale gegeben.

Die Bilddaten werden in dem neuen System nicht mehr in einer Datenbank gehalten, sondern in einem Filesystem. Zum Migrationszeitpunkt erfolgte ein Einspeichern der vorliegenden Bilder. Ein neues Bildlieferverfahren wurde in der ersten Jahreshälfte 2003 konzipiert und implementiert. Dadurch bedingt kam es zu einem mehrmonatigen Stillstand beim Einspeichern der neu digitalisierten Images.¹⁹

Es stehen noch die Programmläufe zur automatischen Verknüpfung der Drucker-/Verleger-Normdatei und der Personennamendatei mit den Titelsätzen aus.²⁰ Voll verknüpfte Normdateien werden sowohl für die Recherche (automatische Suche mit allen Verweisungsformen) als auch für das Bearbeiten der Titelaufnahmen durch die Katalogisierer (u. a. bei Änderungen der Ansetzungen) große Vorteile bringen.²¹

Ende 2002 befanden sich nach Abschluss der Migration 187.200 Titel (inkl. Bände von mehrbändigen Werken) mit insgesamt 330.400 Exemplaren in der neuen VD-17-Datenbank. Mit den Titeln waren 442.320 Bilder von Schlüsselseiten verknüpft. Hinzu kamen 61.900 Datensätze in der Personennamendatei und 4.500 Datensätze in der Drucker-/Verlegerdatei.

WWW-Datenbank

Bilddaten im Filesystem

**Abschluss der Migration
Ende 2002**

FAZIT

Bei einem Treffen der Leiter der VD-17-Arbeitsstellen im Dezember 2002 in Dresden wurde eine (Zwischen-) Bilanz zur Migration gezogen. Alle anwesenden Teilnehmer der Projektbibliotheken und die Vertreter des GBV stimmten darin überein, dass die Migration schnell und erfolgreich verlaufen ist und die neue Datenbank viele Vorteile gegenüber dem alten System bietet. Insbesondere der Wegfall des bisherigen zweistufigen Verfahrens mit den Datenbanken Allegro und OMNIS und des damit verbundenen Zeitaufwands beim Datentransport sowie die wesentlich höhere Aktualität der Datenbank wurden herausgestellt. Wie sich auch danach weiter bestätigte, steht mit PICA ein stabiles bibliothekarisches Standardsystem zur Verfügung. Die bessere Möglichkeit zur Einbindung in Portalsysteme und die modernere und komfortablere WWW-Benutzeroberfläche sind Vorteile, die die Benutzer schätzen werden.

Der notwendige Aufwand auf Seiten des GBV war allerdings zunächst unterschätzt worden. Hierfür gab es mannigfache Ursachen: Zum einen war der Mehraufwand darin begründet, dass u. a. für den Produktionsnummerngenerator und für die Anzeige der Images neue Techniken zum Einsatz kamen. Zum anderen lag die Ursache hierfür in den Spezifika der VD-17-Daten, die dazu führten, dass die Programmierung der Konvertierung und die Parametrisierung der Indizierung und der Anzeige weitaus komplexer wurden als erwartet. Und schließlich führte die wachsende Vertrautheit der VD-17-Katalogisierer mit dem PICA-System nach Aufnahme der Katalogisierung im November und Dezember 2002 zu weiteren Nachbesserungs- und auch Ergänzungswünschen, die von der Verbundzentrale im Rahmen der Möglichkeiten realisiert wurden.

Auf der Katalogisierungsseite war anfangs ein höherer Zeitaufwand bei der täglichen Arbeit deutlich zu spüren, da die aus der alten VD-17-Datenbank über Spielten Titelsätze aufgrund der unterschiedlichen Datenstruktur unter OMNIS und PICA und teilweise durch Konvertierungsfehler eine gewisse Nacharbeit erforderten. Diese konnte durch die bereits erwähnten maschinellen Korrekturläufe der Verbundzentrale einerseits und die Entwicklung von hilfreichen Makros und Skripts andererseits stark verringert werden. Auch der Umgang mit den Personennamensätzen und die Individualisierung der Personennamen erfordern einen gewissen Zeitaufwand, der sich aber langfristig durch eine spürbare Qualitätsanhebung der Datenbank bezahlt machen wird.

Zusammenfassend lässt sich sagen, dass die Migration der VD-17-Datenbank sehr positiv verlaufen ist.

Für die neue VD-17-Datenbank unter PICA sind ein regelmäßiger Support und eine kontinuierliche Weiterentwicklung gewährleistet. Mit der Verbundzentrale des GBV konnte ein guter »Gastgeber« (host) gefunden werden.

¹ www.vd17.de

² Die Portierung wurde durch die HAB Wolfenbüttel geleistet.

³ »Verzeichnis der im deutschen Sprachbereich erschienenen Drucke des XVI. Jahrhunderts«. Informationen zum VD 16 findet man unter www.vd16.de.

⁴ Dies gilt für Abteilung I (Verfasser, Körperschaften, Anonyma). Die Abteilungen II (Register der Herausgeber, Kommentatoren, Übersetzer und Literarischen Beiträger) und III (Register der Druckorte, Drucker, Verleger und Erscheinungsjahre) wurden aber bereits EDV-gestützt erstellt. Rudimentäre Titelaufnahmen (inklusive der Besitzstände) wurden in eine elektronische Datenbank eingebracht. Seit Abschluss des Grundprojekts werden neue Titel und weitere Besitzstände in der Datenbank komplett erfasst. Die Konversion des Grundwerks (Abteilung I) zur Vervollständigung der Titelaufnahmen (insbesondere mit den Sachtiteln) steht aber noch aus.

⁵ Das Datenformat wurde projektspezifisch gegenüber MAB1 leicht erweitert, um z. B. Zensoren, Widmungsempfänger sowie in normierter Form Verleger, Drucker und Druckorte erfassen zu können.

⁶ Zusätzlich war noch eine eigene Bildmanagement-Komponente geplant. Sie sollte es den Katalogisierern ermöglichen, Änderungen in den offline erzeugten und gefüllten Bilddatenbanken vorzunehmen, d. h. Bilder nachträglich einzufügen, zu ersetzen, umzuhängen oder zu löschen. Dieses Modul kam allerdings über das Teststadium nicht hinaus.

⁷ Verbundkatalog maschinenlesbarer Katalogdaten des Deutschen Bibliotheksinstituts

⁸ Zu Projektbeginn existierte die PND noch nicht als konsolidierte Normdatei mit unveränderlichen ID-Nummern.

⁹ In der alten Datenbank war die VD-17-Nummer identisch mit der Produktionsnummer.

¹⁰ in der Form eines sog. Text-BLOB

¹¹ Mittlerweile arbeiteten bis zu 30 Katalogisierer gleichzeitig in der Datenbank.

¹² Client-Software WIN-IBW mit Modifikationen für den VD-17-Einsatz

¹³ Das PICA-System vergibt beim Anlegen einer Titelaufnahme automatisch eine PICA-Produktionsnummer, die in Feld PICA 0100 abgelegt wird.

¹⁴ Die Struktur der VD-17-Nummern ist folgende: Die Nummer beginnt mit dem ein- bis dreistelligen Bibliotheksfernleihsigel, gefolgt von einem Doppelpunkt und einer 6-stelligen laufenden Nummer. Angehängt ist schließlich noch ein Prüfbuchstabe. (Beispiel: 12:123456X).

¹⁵ Im Web-OPAC ist die gedoppelte Fußnote in der bibliographischen Ebene nicht sichtbar. Sie dient nur Indizierungs- und Recherchezwecken.

¹⁶ Vgl.: www.gbv.de/du/katricht/inhalt.shtml

¹⁷ Die bisherige OMNIS-Datenbank steht seither noch als Recherche- und Kontrollinstrument zur Verfügung. Sie wird voraussichtlich Ende 2003 ganz abgeschaltet.

¹⁸ Die Schulungen wurden in Göttingen und Halle durchgeführt. Vorbereitet wurde sie von Wolfenbüttel, Halle und Gotha.

¹⁹ Die Migration der Bilddaten rief erneut das aus der Anfangsphase des VD-17-Projekts resultierende und noch nicht abschließend gelöste Problem mit der nennenswerten Zahl fehlender, falsch indizierter und korrupter Bilder in Erinnerung.

²⁰ Diese sind für Herbst 2003 geplant.

²¹ Die Personennamendatei des VD 17 wird über einen wöchentlichen Änderungsdienst mit der überregionalen Normdatei PND abgeglichen.

DER VERFASSER

Werner Holbach ist Leiter des Referats »Koordination Erschließung« und Leiter des Sachgebiets »VD 17« an der Bayerischen Staatsbibliothek, Ludwigstraße 16, 80539 München
holbach@bsb-muenchen.de