

Aktuelles aus der Digital Library

Christof Mainberger, BSZ Konstanz

Digitales Bibliotheksgut bildet neben dem Verbundsystem und Lokalsystemen schon seit einigen Jahren einen der Schwerpunkte des *Bibliotheksservice-Zentrum Baden-Württemberg (BSZ)*. Dazu wurden in einer Reihe von Projekten unterschiedliche Gesichtspunkte dieser vergleichsweise neuen Medien berücksichtigt. Viele dieser Projekte sind mittlerweile abgeschlossen, einige in einen regelrechten Routinebetrieb übergegangen.

Video- und Audiofiles, aber auch Image- und Textdateien stellen zunächst durch ihre technische Form spezielle Anforderungen an ihre Erzeugung, Aufbewahrung und Nutzung. Daran schließt sich die Entwicklung geeigneter Verfahren und Hilfsmittel zur Verzeichnung und Erschließung an. Spezielle Suchmaschinen und Austauschprotokolle ermöglichen ein adäquates Retrieval elektronischer Ressourcen und ihre Distribution. Ein eigenes Feld stellt der Einsatz von multimedialen Lehr- und Lernmaterialien im Hochschulunterricht dar. Die technischen Eigenschaften und Möglichkeiten führen darüber hinaus zu anderen inhaltlichen Strukturen als bei "konventioneller" Literatur und schließlich zu einer andersartigen rechtlichen Verortung dieser Bestände.

Zu allen diesen Themen war das BSZ tätig, meist in Kooperationen mit Partnern wie z.B. den OPUS-Anwendern oder der *DLmeta-Initiative*. Im Mittelpunkt dieses Engagements steht der Virtuelle Medienserver, der die Metadaten der dezentral vorgehaltenen Objekte enthält, diese über Hyperlinks erreichen kann und der mit der Verbunddatenbank synchronisiert ist.

Die "digitale" Bibliotheksarbeit orientiert sich dabei an den Methoden und Prinzipien der "analogen" Bibliotheksarbeit, passt diese teils den neuen, digitalen Möglichkeiten an, insbesondere der Online-Zugänglichkeit, vermeidet aber Brüche in den Nachweisinstrumenten.

Im Folgenden soll dies an vier zentralen Aspekten deutlich gemacht werden, die Teil jeder Bibliotheksarbeit sind und entsprechend in aktuellen Projekten der Digital Library im BSZ ihren Niederschlag finden: Recherche- und Zugangsmöglichkeiten oder "Portale", Inhalte und Medien oder "Content", Regelwerke und Formate oder "Metadaten", Sprachverwendung oder "Normvokabular und Klassifikationen". Illustriert werden diese Themen anhand aktueller Projekte, zunächst die Sprachverwendung anhand des BAM-Portals:

Das BAM-Portal wird in einem DFG-Projekt in Kooperation des BSZ mit der Landesarchivdirektion Baden-Württemberg und dem Landesmuseum für Technik und Arbeit entwickelt. Es zielt darauf ab, in Bibliotheken, Archiven und Museen vorhandene digitale Bestände unter einer einheitlichen Oberfläche übers World Wide Web zugänglich zu machen. Eine Recherche im BAM-Portal führt auf eine fachübergreifende Trefferliste, in der jeder Treffer über Internetlinks mit einer ausführlichen, herkunftsgerechten Beschreibung verknüpft ist. Von dort ist gegebenenfalls ein zugehöriges Digitalisat bzw. eine multimediale Veranschaulichung erreichbar.

Da übliche Suchaspekte, wie der Autor für Literatur oder die Provenienz für das Archivalien im gemeinsamen Kontext nicht fachübergreifende Resultate ergeben, treten hier themenbezogene Recherchen in den Vordergrund. Daher widmen wir im BAM-Portal der thematischen Erschließung der verschiedenen Datenbestände die größte Aufmerksamkeit.

Dabei stellt sich die Schwierigkeit heraus, dass die Sammlungen im Bibliotheks-, Archiv- und Museumswesen unterschiedliche Ursprünge und unterschiedliche Ziele besitzen und entsprechend nach unterschiedlichen Prinzipien thematisch erschlossen sind.

Bibliothekare betrachten Ausgaben einzelner Werke als in sich geschlossene, in der Regel unverbundene, "atomare" Einheiten. Diese werden durch Schlagwortketten aus Normvokabular sachlich erschlossen.

Die Ordnung von Archivgut spiegelt demgegenüber Verwaltungshandeln wider. Entsprechend stellt sich Archivgut als gegliederte Systematik ineinander enthaltener Bestände und Teilbestände dar, deren thematische Verortung durch die Stelle in der Systematik markiert ist.

Museen stellen teils einzelne Exponate, teils ganze Sammlungen in den Mittelpunkt. Ihre Methode ist die dokumentarische Erschließung, die mit Schlagworten arbeitet, sich aber auch auf umfangreiche Beschreibungen und Abstracts stützt.

Nachdem eine auf Sammlungen orientierte Sicht für Literatur nicht sinnvoll vorzunehmen, eine "atomare" Betrachtungsweise aber grundsätzlich auch auf Archiv- und Museumsgut anwendbar ist, werden im BAM-Portal die einzelnen Ressourcen als eigenständige Einheiten zu Grunde gelegt. Entsprechend folgt auch die thematische Erschließung dem bibliothekarischen Vorbild, so dass auch für Archivalien und Exponate eine Beschlagwortung vorgenommen wird.

Eine Harmonisierung der Erschließungsmethode führt dann zu übergreifenden Ergebnissen, wenn auch in allen zusammengeführten Beständen die gleichen Inhalte mit den gleichen Begriffen bezeichnet werden, wenn also ein gemeinsames normiertes Vokabular Verwendung findet. Für das BAM-Portal wurde vereinbart, dazu die Schlagwortnormdatei (SWD) der Deutschen Bibliothek einzusetzen, nicht nur weil die Literatur und ein Teil der Museumssammlungen mit diesen Schlagworten bearbeitet sind, sondern auch weil keine annähernd ausgearbeiteten Alternativen aus dem Archiv- oder Museumsbereich existieren.

Dem Einsatz der SWD, die bislang nur den Katalogisierungswerkzeugen der Bibliothekare vorbehalten war, muss außerhalb dieses Kontexts allerdings erst der Weg bereitet werden:

Ein erster Schritt bildet dazu die OSWD, also SWD im WWW. Dieses Werkzeug wurde vom Bibliotheksservice-Zentrum für die Vergabe von Metadaten für online-verfügbare Dokumente geschaffen. Autoren von Hochschulschriften können damit ihre Publikationen mit regelgerechten Deskriptoren versehen, die aus einer Weboberfläche ermittelt werden. Im BAM-Projekt kommt diese Webapplikation nun auch Archivaren und Museologen zugute, die hier nach passenden Schlagworten für ihr Material suchen können.

In etlichen Fällen führte diese Suche allerdings bislang nicht zu geeigneten Ergebnissen: Die SWD, für bibliographische Bedürfnisse geschaffen, enthielt Begriffe wie Schuhbürste oder Aschenbecher nicht, die für die Beschreibung entsprechender Exponate oder Archivalien unverzichtbar sind. Um diese Defizite auszugleichen, vereinbarte das BSZ mit der Deutschen Bibliothek ein Verfahren, nach dem Museen und Archive eigene Deskriptoren zu den Normdateien beitragen können, und erweiterte die SWD im WWW um eine Eingabeoberfläche. Im Beitrag von Frau Dr. Hesse von den Reich-Engelhorn-Museen beim BSZ-Kolloquium werden erste positive Erfahrungen mit dieser Anwendung berichtet, die nach Abschluss der Projektphase im Regelbetrieb weitergeführt wird.

Neben der thematischen Suche erfordert ein weiteres Vorhaben im BAM-Portal eine gemeinsame, geregelte Sprachverwendung: Den Nutzern soll über eine Verzeichnisstruktur ein navigierendes Recherchieren ermöglicht werden. Erforderlich ist dazu zunächst eine Klassifikation, die alle Themen umfasst und dann eine Einordnung der Ressourcen in diese Klassifikation. Da dies manuell nicht realisierbar ist, müssen automatische Wege zur Einordnung gefunden werden.

Einen Ansatz dazu bietet die Regensburger Verbundklassifikation, die ausreichend universell ist und deren Registerinträge an der SWD abgeglichen sind. Nachdem die Bestände aus den Archiven und Museen mit SWD-Schlagworten erschlossen sind, gelingt über die Relationierung der Verbundklassifikation zur SWD die Einordnung der Objektbeschreibungen.

Ein Einwand gegen die Überarbeitung des Archiv- und Museumsguts mit geregelter Vokabular für das BAM-Portal stellt der hohe Aufwand dar, der es erlaubt, nur unbedeutende Bruchteile der beträchtlichen Bestände einzubinden. Daher werden im Bibliotheksservice-Zentrum automatische Indizierungsanwendungen getestet, die die computergestützte Zuordnung von Deskriptoren zu Katalogisaten ermöglichen.

Diese Verfahren unterziehen die Textbestandteile einer Objektbeschreibung zunächst einer linguistischen Analyse, durch die das Wortmaterial auf grammatikalische Grundformen zurückgeführt wird, Stoppwörter gekennzeichnet und Komposita getrennt werden. Eine nachfolgende statistische Analyse ermittelt dann Begriffe, die für den Text typisch sind. Abschließend werden diese Begriffe auf Normvokabular abgebildet, wodurch sich die gesuchten Deskriptoren ergeben. Erfahrungen aus den Tests zeigen, dass diese Methode brauchbare Ergebnisse erzielt, mindestens aber eine nützliche Grundlage für eine intellektuelle Beschlagwortung darstellt.

Ein letzter Aspekt einer geregelten Sprachverwendung betrifft den Nutzer: Wenn er z.B. in einem Webangebot wie dem BAM-Portal die ihm geläufige Alltagssprache einsetzen will, müssen seine Anfragen – gleichfalls automatisch – auf das Normvokabular abgebildet werden, das der Datenbasis zugrunde liegt. Sucht er nach "Süßwaren", sollte das System diesen Begriff in seiner Anfrage durch die korrekte SWD-Ansetzung "Süßware" ersetzen können, sowie deren Unterbegriffe "Dauerbackware", "Feingebäck", "Gummibärchen" etc. einfügen, so dass die vom Nutzer erwarteten Resultate vollständig ermittelt werden können. Da eine solche Umsetzung auch für andere Rechercheinstrumente nützlich ist, wird sie in einem vom BAM-Portal unabhängigen Webservice implementiert.

Für die Sprachverwendung adaptiert die "digitale" Bibliothek also die Normdateien, die die "analoge" Bibliothek bereitgestellt hat, wendet sie allerdings auf neue Bereiche wie Archiv- und Museumsgut an, passt lediglich Werkzeuge z.B. an die Internetumgebung an und sucht insbesondere Wege, um mit der Fülle nun anfallenden digitalen Materials umgehen zu lernen.

Anders stellt sich die Situation bei den Metadaten und Datenformaten dar. Hier sind folgende Kriterien zu nennen:

- Metadaten für die digitale Bibliothek müssen einfach sein.
- Metadaten für die digitale Bibliothek müssen der Komplexität und der technischen Beschaffenheit multimedialer Objekte gerecht werden.
- Metadaten für die digitale Bibliothek sollten international sein.
- Metadaten für die digitale Bibliothek werden als XML-Anwendung formuliert.

Diese Kriterien führen auf Vorschläge, wie z.B. Dublin Core, die wesentlich offener und weniger ausgearbeitet sind als etwa das Regelwerk RAK oder das Format MAB, damit aber auch weniger verbindlich.

Ein typischer Einsatz solcher einfachsten Metadaten stellt das Projekt "Elektronische Semesterapparate" dar, das das Bibliotheksservice-Zentrum gemeinsam mit der Bibliothek der Universität Konstanz durchführt. Dabei handelt es sich im Kern um ein Content-Management-System, das es Dozenten der Universität Konstanz über eine Webapplikation erlaubt, digitales Lehrmaterial in einer Datenbank zu verzeichnen, auf einem Server abzulegen und in einer hierarchischen Struktur den Studierenden zu präsentieren.

Um die Benutzung der Applikation attraktiv zu machen und den Aufwand möglichst gering zu halten, werden den Dozenten bei der Verzeichnung von Lehrmaterial nur die nötigsten Angaben abverlangt, die auch direkt zur Präsentation und Verlinkung des Lehrmaterials benötigt werden. In dem vom Bibliotheksservice-Zentrum in Zusammenarbeit mit dem Zentrum für Datenverarbeitung Tübingen (ZDV) und dem Lehrstuhl für Management und Organisation von Informationssystemen Ulm (LOMI) entwickelten Metadatenformat für die digitale Bibliothek DLmeta sieht ein solcher Datensatz folgendermaßen aus:

```
<DLmeta Version="2002">
  <Object ObjectId="123" Origin="Esem">
    <DC>
      <Title Role="TitleMain" Language="ger">
        Kapitel 3: Methoden des E-Learnings
      </Title>
      <Description Role="DescriptionMain">
        Der Einsatz von computergestützten ...
      </Description>
      <Identifizier Scheme="URL">
        http://www.ub.uni-konstanz.de/esem/doc/abc/kap3-pdf
      </Identifizier>
    </DC>
  </Object>
</DLmeta>
```

Eine wesentliches Merkmal des Projekts "Elektronische Semesterapparate" liegt in dessen Einbettung in die übrige IT-Infrastruktur der Bibliothek der Universität Konstanz. So enthalten die "Elektronischen Semesterapparate" keine eigene Komponente zur Authentifizierung seiner Benutzer, sondern greifen dazu auf die Benutzerdatenbank der Universitätsbibliothek zurück. Und die Applikation fragt die Ausleihdatenbank der Bibliothek ab, um auch die konventionell in Semesterapparaten zur Verfügung gestellte Literatur neben den digitalen Materialien einzubinden.

Auch die Veröffentlichung von elektronischem Lehrmaterial im Sinne einer Hochschulpublikation soll über ein schon vorhandenes, für solche Zwecke vorgesehenes Werkzeug geschehen, über das in der Bibliothek installierte OPUS-System KOPS. Die etablierten Verfahren der Datenlieferung an den Virtuellen Medienserver des BSZ und dessen Synchronisation mit der Verbunddatenbank führen schließlich zum Nachweis des elektronischen Lehrmaterials im Südwestverbund.

Dabei sollen die schon erhobenen, wenn auch spärlichen Metadaten, die die Dozenten in die Formulare der "Elektronischen Semesterapparate" eintragen, über das XML-Austauschformat an den Hochschulpublikationsserver vermittelt und in dessen Eingabemasken übernommen werden. Um eine OPUS-konforme Verzeichnung zu erstellen, muss sie der Autor lediglich um die Felder – im Wesentlichen Schlagworte – ergänzen, die zur regelgerechten Publikation obligatorisch

sind. Der oben dargestellte Metadatensatz würde nach einer solchen "Hochkatalogisierung" um Subject-Elemente ergänzt sein (die Scheme-Angabe in den Subject-Elementen markiert die oben erwähnte SWD-Verwendung in den OPUS-Systemen):

```
<DLmeta Version="2002">
  <Object ObjectId="123" Origin="Esem">
    <DC>
      <Title Role="TitleMain" Language="ger">
        Kapitel 3: Methoden des E-Learnings
      </Title>
      <Subject Scheme="SWD">
        Pädagogik
      </Subject>
      <Subject Scheme="SWD">
        E-Learning
      </Subject>
      <Subject Scheme="SWD">
        Computer
      </Subject>
      <Description Role="DescriptionMain">
        Der Einsatz von computergestützten ...
      </Description>
      <Identifier Scheme="URL">
        http://www.ub.uni-konstanz.de/esem/doc/abc/kap3-pdf
      </Identifier>
    </DC>
  </Object>
</DLmeta>
```

Die Berücksichtigung unterschiedlicher Tiefen der Erschließung und die häufige Transformation sind Charakteristika von Metadaten in einer digitalen Bibliothek. Metadaten müssen sich dabei einerseits der – oft unzureichenden – Kompetenz derer anpassen, die die Ressourcen erschließen sollen. Andererseits sollen sie in der Lage sein, auch komplexe technische Eigenschaften von digitalen Medien detailliert zu beschreiben, etwa um geeignete Such- und Navigationsvorgänge auf ihnen zu unterstützen.

Mit der XML-Anwendung DLmeta hat das Bibliotheksservice-Zentrum gemeinsam mit dem Zentrum für Datenverarbeitung in Tübingen und dem Lehrstuhl für die Organisation und Management von Informationssystemen in Ulm ein Datenmodell und -austauschformat entwickelt, das dieser Forderung nach Skalierbarkeit in der semantischen Komplexität gerecht wird.

Es teilt die Metadatenelemente in vier Gruppen ein: Die erste Gruppe enthält im Wesentlichen die Elemente des Dublin Core Metadata Element Set; die zweite Gruppe bietet einen Erweiterungsmechanismus, der anwendungsspezifische und lokale Ergänzungen erlaubt. Die dritte Gruppe beschreibt die technische Struktur der Ressourcen und ermöglicht es, digitale Inhalte in die Beschreibungsdaten einzubetten. Die vierte Gruppe enthält schließlich Meta-Metadaten, also Daten, die die Metadaten selbst beschreiben und so ihre Verwaltung erleichtern.

Die Formulierung von DLmeta als XML-Anwendung legt den Einsatz von XML-Werkzeugen nahe. Vor allem ist hier der frei verfügbare XSLT-Transformator XALAN zu nennen, der im BSZ zur Transformation von XML-Dokumenten verwendet wird, mit dessen Hilfe aber auch z.B. MAB-Diskettenformat-Dateien verarbeitet oder Ladefiles für Datenbanken erstellt werden können.

Diese Technologie kann z.B. für die Tübinger Schallplattensammlung angewandt werden. Ausgangspunkt sind hier im Südwestverbund katalogisierte Schallplatten. Diese Katalogisate werden per KATWIN-Download heruntergeladen, via XSLT in DLMeta transformiert, mit Hilfe des im ZDV Tübingen entwickelten DLMeta-Editors ergänzt und dann über den Virtuellen Medienserver in die Verbunddatenbank zurücktransportiert.

An diesem Anwendungsszenario wird besonders gut deutlich, dass bei allen syntaktischen, semantischen und konzeptionellen Differenzen von Metadaten in der digitalen Bibliothek immer wieder ihr Ausgangspunkt oder Ziel die Verbunddatenbank und damit ihre Konkordanz zu etablierten Formaten und Regelwerken ist.

Der dritte Abschnitt dieses Beitrags behandelt die Aktivitäten des Bibliotheksservice-Zentrum bezüglich digitaler Inhalte und Medien, sogenannter Contents. Digitale Objekte werden im Südwestverbund in der Regel im dezentralen Depot, das heißt bei den Einrichtungen gehalten, die sie erzeugen. Für die Volltexte und Hochschulpublikationen ist hier vor allem das Netzwerk von Hochschulschriftenservern zu nennen, für Videos z.B. der Tübinger Internet Multimedia Server (TIMMS) oder das Digitale Video Archiv (DIVA) in Karlsruhe.

Das BSZ hat die Aufgabe, Beschreibungsdaten zu den Objekten im Virtuellen Medienserver zusammenzuführen und in die Verbunddatenbank zu vermitteln. Aber die Übergänge zwischen digitalen Objekten und Verzeichnungsdaten können fließend sein: Rezensionen, Abstracts und Klappentexte, die in der SWB+ Datenbank (www.bsz-bw.de/rekla) vorgehalten werden, gehören zwar inhaltlich zu den Beschreibungsdaten, unterscheiden sich aber technisch nicht von digitalen Volltexten.

Weitere Aktivitäten des BSZ zu digitalen Objekten betreffen eher kleinere Einrichtungen, wie Archive, Museen, Berufsakademien oder Fachhochschulen, für die das BSZ Objekt- und Applikation-Server zur Verfügung stellt. Ein Beispiel dafür ist die Interessengemeinschaft Ariadne Baden-Württemberg (IGA). Ariadne ist ein europäisches Netzwerk zur Erstellung, Verzeichnung, Präsentation und zum Austausch von Lehr- und Lernmaterial. Das BSZ betreibt einen Ariadne Server, der sich im Rahmen der IGA insbesondere an Fachhochschulen richtet.

Die IGA hat zum Ziel, in Baden Württemberg eine dezentralisierte Infrastruktur von Ariadne-Servern einzurichten, die einen homogenen Zugang zu Lehr- und Lernmaterial, insbesondere der Virtuellen Hochschule erlaubt, eine adäquate und zur Verbunddatenbank konkordante Verzeichnung dieses Materials gewährleistet und eine qualifizierte Suche über diese Bestände ermöglicht. Neben dem BSZ sind in der Gemeinschaft die Universitäten Karlsruhe, Freiburg, Stuttgart, Tübingen die Virtuelle Hochschule Oberrhein (VIROR) sowie der Verbund Virtuelles Labor (VVL) vertreten.

Der letzte Abschnitt dieses Beitrags beschäftigt sich mit Recherche- und Zugangsmöglichkeiten, so genannten "Portalen". Bislang gibt es noch kein umfassendes Bibliotheksportal in Baden-Württemberg, obwohl einzelne Komponenten dazu vorhanden sind. Ein Beispiel ist hier der "Suchdienst für wissenschaftliche Bibliotheken (SWIB)" der über den Inhalt des Virtuellen Medienserver hinaus weitere online verfügbare Materialien zusammenführt.

Eine weitere Initiative auf diesem Gebiet könnte die Etablierung eines Portals für Fachhochschulen, Berufsakademien und Museen darstellen, die nicht über genügend Kapazität verfügen, um selbst einen solchen integrierten Service für ihre Nutzer bereitzustellen.

Ein eigentliches Portal stellt aber selbstverständlich auch das BAM-Portal da, indem es eine Reihe typischer Portalfunctionalitäten realisiert: Es bietet einen homogenen, übergreifenden Zugriff zu Ressourcen aus Bibliotheken, Archiven und Museen. Es fragt Bestände, die über das Internet erreichbar sind, direkt ab und sammelt bei Beständen, die nicht online verfügbar sind, die Meta- und Objektdaten in einen zentralen Index. Dies ermöglicht, dass auch einrichtungsspezifische Profile und Sichten auf den BAM-Index eingerichtet werden.

Umgekehrt unterstützt das BSZ Portale von Dritten, indem es seine eigenen Bestände möglichst ungehindert zugänglich macht. Die wichtigsten Schnittstellen hierfür sind der WebPAC unter Z39.50 und dem HTTP-Protokoll und der Suchdienst wissenschaftlicher Bibliotheken SWIB, der seinerseits von Dritten in eine Reihe von Recherchewerkzeuge eingebunden wurde.

Eine weitere Distributionsmöglichkeit für Daten bildet die OAI-Schnittstelle zum Virtuellen Medienserver, die bei der Open Archive Initiative als Daten-Provider registriert ist. Über diese Schnittstelle können die Metadaten komplett oder teilweise in ein anderes System geliefert werden, das sie nachnutzt und seinerseits anbietet. Seit der Auslieferung der OPUS 2.0 Version wird das OAI-Protokoll auch genutzt, um die Metadaten aus dem OPUS-Installationen in den Virtuellen Medienserver zu übertragen.

Ein letzter Aspekt betrifft die Erreichbarkeit und Identifizierung von digitalen Objekten. Die Idee ist, dies nicht wie bislang über die URL zu gewährleisten, die sich erfahrungsgemäß oft ändert. Stattdessen soll jede Ressource einen unveränderbaren Namen, einen sogenannter Unified Resource Name (URN), erhalten, der in einem Name-Server auf die aktuelle URL abgebildet wird.

Der Vorteil ist dass die URLs nur noch im Name-Server aktualisiert werden müssen, während alle aus URNs gebildeten Links auf eine Ressource korrekt bleiben. Das Bibliotheksservice-Zentrum hat mit der Deutschen Bibliothek das Verfahren vereinbart, nach dem URNs im Verbund verwaltet werden, und hat die erforderlichen Werkzeuge bereitgestellt, um URNs zu vergeben, zu registrieren, aktuell zu halten und aufzulösen.

Damit endet dieser Beitrag. Anhand der Themen Sprachverwendung, Metadaten, Content und Portale wurden die DL-Aktivitäten des BSZ im vergangenen Jahr exemplarisch dargestellt. Dabei muss betont werden, dass fast jedes Projekt jeweils alle vier Aspekte berührt, die insofern als ein roter Faden die Arbeiten zur Digital Library durchziehen.

Christof Mainberger
Bibliotheksservice-Zentrum Baden-Württemberg
Universität Konstanz
D-78457 Konstanz
Tel.: +49 (0) 75 31/88 44 68
E-Mail: christof.mainberger@bsz-bw.de