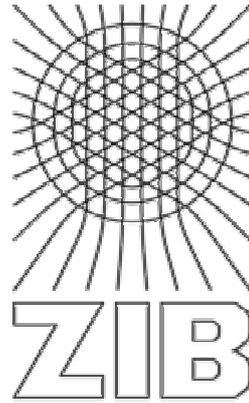

Konrad-Zuse-Zentrum
für Informationstechnik Berlin



Takustraße 7
D-14195 Berlin-Dahlem
Germany

IVAN BOEV, LAVINIA HODOROABA, ANDRES IMHOF

Der KOBV-Metadaten-Austausch-Parser (KMA-Parser): Ein Werkzeug für Metadata-Sharing

Gefördert
von der Senatsverwaltung für Wissenschaft, Forschung und Kultur des Landes Berlin,
vom Ministerium für Wissenschaft, Forschung und Kultur des Landes Brandenburg
und von den Mitgliedsbibliotheken des KOBV

Der KOBV-Metadaten-Austausch-Parser (KMA-Parser): Ein Werkzeug für Metadata-Sharing

Ivan Boev, Lavinia Hodoroaba, Andres Imhof

Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)

ZIB Report 05-29 (Mai 2005)

Abstract

Der Kooperative Bibliotheksverbund Berlin-Brandenburg (KOBV) betreibt für das KOBV-Portal einen Austausch von Beschreibungen zu elektronischen Ressourcen (Metadaten), der zwischen verschiedenen Informationsportalen in der Region durchgeführt wird. Die unterschiedlichen Informationsportale verwenden verschiedene Metadaten-Schemata und Austausch-Formate für ihre Ressourcebeschreibungen. Um die Heterogenität der Metadaten zu überwinden, wurde der KOBV-Metadaten-Austausch-Parser (KMA-Parser) entwickelt. Andere Metadaten-Schemata werden auf das KOBV-Metadaten-Schema abgebildet. Der KMA-Parser führt gegebenenfalls eine Format-Transformation durch, konvertiert die Inhalte einzelner Elemente über Konkordanzen und erzeugt neue Metadaten-Elemente aus vorhandenen Feldern. Er validiert den Inhalt auf Vollständigkeit und steuert den Austausch der Metadaten zwischen den Portalen. Der Umwandlungsprozess funktioniert jedoch nicht nur in die Richtung des KOBV-Portals, sondern auch in die entgegengesetzte Richtung. Der Artikel beschreibt die einzelnen Vorgänge im KMA-Parser und schildert die Erfahrungen im Umgang mit der Heterogenität von Metadaten. Die gewonnenen Erfahrungen verweisen auf grundlegende Perspektiven in der universellen Zusammenarbeit von Informationsanbietern und -Providern.

The Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV) has built a framework for a bi-directional exchange workflow of electronic resources' descriptions (metadata) between the KOBV Portal and other Information Portals in the region. The Information Portals use different exchange formats, metadata schemata and controlled vocabularies for their descriptions of resources. In order to overcome this metadata heterogeneity, an application, the KOBV Metadata Exchange Parser (KMA-Parser), has been developed. The KMA-Parser maps the local portals' metadata schemata into the metadata schema of the KOBV Portal. If necessary, it transforms the exchange format, converts contents of individual elements by means of concordances and produces new metadata elements on the basis of existent elements. It checks elements' contents on completeness and controls the metadata exchange between the portals. However, the transformation process takes place not only towards the KOBV Portal, but also vice versa. The article describes the individual processes in the KMA Parser and depicts the experiences in handling the metadata's heterogeneity. The experiences gathered give an idea of the prospects for a universal cooperation between information suppliers and providers.

Keywords: KOBV-Portal, Metadata-Sharing, Metadaten-Austausch, Vernetzung von Portalen

CR: H.OK, 4.0

Inhalt

Einleitung.....	3
Der Einsatzort.....	4
Die Aufgaben	5
<i>Format-Transformation.....</i>	<i>5</i>
<i>Konvertierung über Konkordanzen</i>	<i>6</i>
<i>Bildung eines neuen Metadaten-Elementes.....</i>	<i>6</i>
<i>Synchronisation</i>	<i>6</i>
<i>Kontrolle und Auswahl.....</i>	<i>6</i>
<i>Wechselseitiger Metadatenaustausch.....</i>	<i>7</i>
<i>Aufbau und Handhabung des KMA-Parsers</i>	<i>7</i>
Die Konfiguration	8
Konkordanzen	10
Erfahrungen und Ergebnisse	11
<i>Unterstützung von lokalen Benennungen und Standards gleichermaßen.....</i>	<i>11</i>
<i>Qualitätskontrolle der Ressourcebeschreibungen.....</i>	<i>11</i>
<i>Kontrolle über Ressourcenauswahl nach selbst bestimmten Kriterien</i>	<i>12</i>
<i>Empfehlung zur Berücksichtigung von Standards bei allen Partnern.....</i>	<i>12</i>
Fazit.....	12

Einleitung

Bibliotheken betreiben einen hohen finanziellen Aufwand, um für ihre NutzerInnen aus Forschung und Lehre lizenzierte Ressourcen¹ bereitzustellen. Diese Ressourcen werden normaler Weise statisch auf den lokalen Webseiten oder in lokalen Portalen aufgeführt. Um sich als NutzerIn in der Region Berlin-Brandenburg einen generellen Überblick über die verfügbaren Ressourcen zu verschaffen, musste man bislang von Bibliotheksseite zu Bibliotheksseite und von Portal zu Portal surfen. Das KOBV-Portal² hat diesen Mangel beseitigt und verzeichnet unter einer Oberfläche alle lizenzierten Ressourcen der Bibliotheken in der Region neben einer qualifizierten Auswahl frei verfügbarer Ressourcen.

Im Jahr 2004 führte die KOBV-Zentrale³ diese Zielsetzung mit dem Projekt „KOBV-Portal II – Metadata-Sharing“ fort, um die Integration von weiteren freien und lizenzierten Ressourcen durch kooperative Erschließung und automatisierten Datenaustausch voranzutreiben. Der KOBV-Metadaten-Austausch-Parser⁴, im Folgenden kurz KMA-Parser genannt, ist neben dem KOBV-Metadaten-Schema⁵, dem Metadata-Tool⁶ und der Redaktion freier Ressourcen⁷ ein Teilprojekt dieses Vorhabens.

Eine wichtige Anforderung an das Projekt „KOBV-Portal II – Metadata-Sharing“ war, dass den Bibliotheken kein Mehraufwand entstehen darf, wenn die Informationen über die lizenzierten Ressourcen in das KOBV-Portal gelangen, denn die Metadaten kommen weiterhin von den Bibliotheken selbst. Hierfür setzt die KOBV-Zentrale das Metadata-Sharing ein. Einmal angelegte Ressourcebeschreibungen sollen mehrfach verwendet werden können.

Geprüft wurde u. a. die Möglichkeit, lokale Webseiten der Bibliotheken über ein Web Crawling auszulesen. Ein Web Crawler⁸ hätte jedoch bei jeder Veränderung der HTML-Seiten der Bibliotheken angepasst werden müssen, was für diesen Zweck zu aufwändig in der fortlaufenden Pflege gewesen wäre. Außerdem haben diese HTML-Seiten nicht alle für das KOBV-Portal obligatorischen Metadaten enthalten. Dass strukturierte Metadaten zu den elektronischen Ressourcen vorliegen, ist aber die Vorbedingung für ein Metadata-Sharing. Aus diesem Grund konnte auch nicht das OAI-Protokoll⁹ für den Metadaten-Austausch eingesetzt werden, weil es keine strukturierten Metadaten gab, die über

¹ Der Begriff Ressource wird in diesem Zusammenhang als dokumentarische Einheit innerhalb der Informationswissenschaften aufgefasst; synonym kann auch Informationsquelle verwendet werden.

² Das KOBV-Portal ist Ende Dezember 2003 in Betrieb gegangen. Informationen über das KOBV-Portal siehe: Lavinia Hodoroaba, Andres Imhof, Monika Kuberek: Das KOBV-Portal, elektronische Ressourcen in Berlin-Brandenburg: Nachweis, parallele Suche und weiterführende Dienste. In: Bibliotheksdienst 38. Jg. (2004) H. 9, S. 1055. – Volltext unter: <ftp://ftp.zib.de/pub/zib-publications/reports/ZR-04-31.pdf>

³ Informationen zum Kooperativen Bibliotheksverbund Berlin-Brandenburg (KOBV) unter: <http://www.kobv.de>

⁴ In diesem Zusammenhang ist ein Parser ein Programm, das Dokumente auf ihre Struktur hin überprüft und verarbeitet. Die Syntax der Metadaten und ihre Werte werden eingelesen und den vorgegebenen Kriterien entsprechend verändert.

⁵ Lavinia Hodoroaba, Andres Imhof: Das KOBV-Metadaten-Schema im KOBV-Portal. ZIB-Report 04-28 (August 2004). – Volltext unter: <ftp://ftp.zib.de/pub/zib-publications/reports/ZR-04-28.pdf>

⁶ Das Metadata-Tool ist ein webbasiertes Eingabe-Tool, das die dezentrale Erfassung von Ressourcebeschreibungen ermöglicht, die zentral gespeichert werden. Bibliotheken, die über kein eigenes lokales Bibliotheksportal verfügen, können über das Metadata-Tool ihre lizenzierten Ressourcen – jeweils auf dem aktuellen Stand – dem KOBV-Portal melden.

⁷ Die Redaktion freier Ressourcen ist durch einen ersten Workshop am 22.09.2004 initiiert worden. Ziel ist es kooperativ, nach einzelnen Fachgebieten geordnet einzelne Fachredakteure zu bestimmen und so auf mehrere Institutionen verteilt die Redaktion zu freien, wissenschaftlich relevanten Ressourcen zu koordinieren.

⁸ Ein Web Crawler ist ein Programm, das andere Webseiten besucht und den Inhalt für eine Indizierung herunterlädt.

⁹ <http://www.openarchives.org/>

eine OAI-Schnittstelle hätten ausgetauscht werden können. Also wurde zunächst nach Wegen gesucht, wie strukturierte Metadaten erschlossen werden können. Neben den wenigen, lokalen Portallösungen, aus denen die Metadaten bezogen werden, wurde für die Mehrzahl an Bibliotheken ohne ein eigenes Portal ein webbasiertes Eingabe-Tool geschaffen, über das strukturierte Metadaten zu elektronischen Ressourcen erfasst werden können.

Die nun aus den Bibliotheken vorliegenden, strukturierten Metadaten müssen jedoch nicht nur transportiert, sondern auch in Teilbereichen bearbeitet werden, da jede Bibliothek in den Metadaten-Elementen individuelle Ansätze zeigt. Der KMA-Parser ist eine in Perl und XML entwickelte Applikation, die die Beschreibungen zu Ressourcen (Metadaten) aus einem Portal universell für eine weitere Verwendung in anderen Portalen umwandelt. Nach konfigurierbaren Parametern werden vorhandene Ressourcenbeschreibungen aufbereitet und quasi wie bei einer Titelaufnahme im Katalog für das jeweilige Portal auf die lokalen Bedingungen angepasst. Auf diese Weise wird mit Hilfe des KMA-Parsers ein Austausch auch von sich unterscheidenden Metadaten-Schemata durchgeführt. Das ermöglicht ein Metadata-Sharing zwischen mehreren Portal-Betreibern in großem Datenumfang. Einmal erarbeitete Informationen können so mehrfach, vielseitig und flexibel eingesetzt werden. Im Folgenden werden die technischen Merkmale, Vorgehensweisen und Erfahrungen im Umgang mit heterogenen Metadaten erläutert.

Der Einsatzort

Die KOBV-Zentrale betreibt für die Region Berlin-Brandenburg das KOBV-Portal. Es baut auf der KOBV-Suchmaschine auf, die seit 1999 als virtueller Verbundkatalog der Region Berlin-Brandenburg genutzt wird. Wie bereits zuvor bei der KOBV-Suchmaschine wird die Portalsoftware MetaLib der Firma Ex Libris eingesetzt. Das erweiterte KOBV-Portal weist zum einen die lizenzierten Ressourcen für jede einzelne Bibliothek in der Region nach, zum anderen können wissenschaftlich relevante Ressourcen, die frei zugänglich sind, recherchiert werden. Beides ist in der Suche aus Nutzersicht nach thematischen Gesichtspunkten, Ressourcotyp und Zugriffsart differenzierbar.

Um die notwendigen Informationen für den Bestandsnachweis der lizenzierten Ressourcen im KOBV-Portal zu beziehen, wird auf bereits bestehende strukturierte Metadaten in den jeweiligen Bibliotheken zurückgegriffen. Denn dort werden sie angelegt und aktuell gehalten. Welche Systeme zum Anlegen strukturierter Metadaten zum Einsatz kommen, ist in Abbildung 1 dargestellt:

- Einige Bibliotheken in der Region beteiligen sich am Konsortialmodell für MetaLib als Portalsoftware, das in der KOBV-Zentrale gehostet wird¹⁰.
- Andere Bibliotheken betreiben ihr lokales Portal mit einer anderen Software. Doch auch Ressourcebeschreibungen der Fremdsysteme können für das KOBV-Portal aufbereitet werden.
- Jedoch besteht bei den wenigsten Bibliotheken bislang die Möglichkeit, strukturierte Metadaten zu ihren elektronischen Ressourcen anzulegen. Für diese wird in Zukunft von der KOBV-Zentrale ein webbasiertes Eingabe-Tool, das so genannte Metadata-Tool, angeboten. Damit können die Bibliotheken ihre lizenzierten Ressourcen pflegen und direkt in das KOBV-Portal einbringen.

Für die freien Ressourcen ist eine verteilte Redaktion in der Region initiiert worden, die zu allen Wissenschaftsgebieten FachredakteurInnen aus den Bibliotheken gewonnen hat, die die Erschließung und Pflege der Ressourcebeschreibungen übernehmen. Die Redaktion wird ebenfalls das Metadata-Tool verwenden, nur dass sein Einsatz jetzt den freien und nicht den lokalen, lizenzierten Ressourcen einzelner Bibliotheken gilt.

Die strukturierten Metadaten kommen also aus verschiedenen Systemen. Da die Metadaten in der Regel ein anderes Format aufweisen, wenn sie nicht aus einem MetaLib-System stammen, wie es zumindest für die Konsortialpartner, das Metadata-Tool und das Redaktionssystem zutrifft, müssen diese aus Fremdsystemen als erstes in das MetaLib-XML-Format umgewandelt werden. Anschließend müssen aber auch die Inhalte einiger Metadaten-Elemente angepasst werden. D. h.

¹⁰ Informationen zum MetaLib/SFX-Konsortialmodell des KOBV unter: http://www.kobv.de/deutsch/content/partner_info/arbeitspapiere/kobv-konsortialportal_modelle.pdf

bestimmte Metadaten-Elemente müssen analog zum KOBV-Metadaten-Schema, das speziell für das KOBV-Portal entwickelt worden ist, konvertiert werden. Hierfür werden u. a. spezielle Konkordanzen zu vier Metadaten-Elementen eingesetzt. Außerdem müssen einige Abgleichungen innerhalb des Ex- und Imports zwischen den lokalen Bibliotheksportalen und dem KOBV-Portal übernommen werden, um beispielsweise zu erkennen, ob eine Ressource im lokalen Portal entfernt wurde oder nicht.

Für diese Aufgaben wurde der KMA-Parser entwickelt. Er ist das Verbindungsstück zwischen den verschiedenen Modulen des KOBV-Portals: den lokalen Bibliotheksportalen, dem Metadata-Tool sowie dem Redaktionssystem auf der einen Seite und dem KOBV-Portal auf der anderen Seite (siehe Abbildung 1).

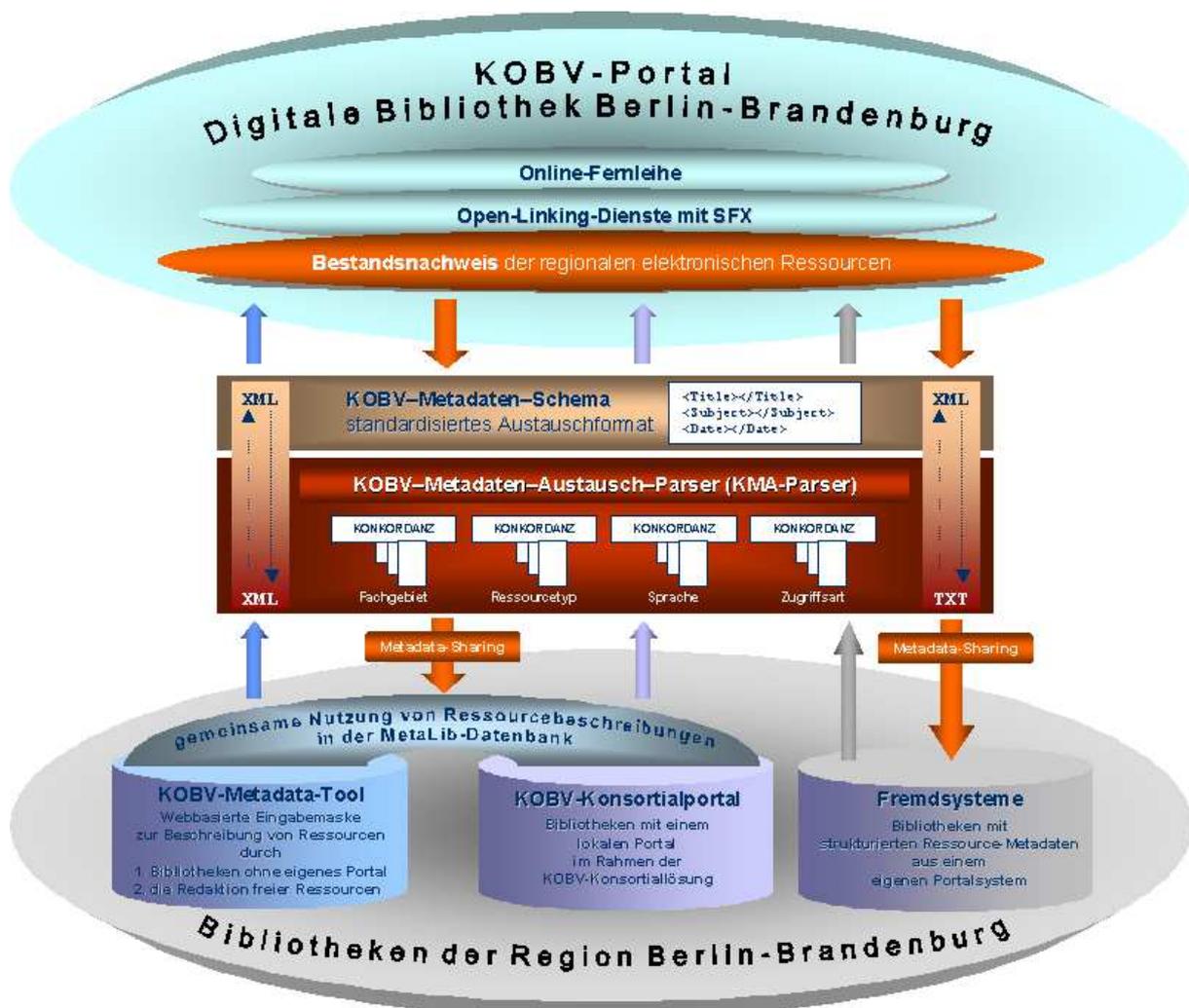


Abbildung 1: Das Modell des KOBV-Portals mit den einzelnen Modulen und Komponenten.

Die Aufgaben

Für das Metadata-Sharing und den Umgang mit unterschiedlichsten Schemata hat der KMA-Parser verschiedene Aufgaben zu erfüllen:

Format-Transformation

Für die Bibliotheken, die nicht MetaLib als Portalsoftware verwenden, muss eine Transformation der gelieferten Metadaten in das Import-Format des KOBV-Portals, nämlich eine Variante von Marc21 in

XML¹¹, vorangestellt werden. Der KMA-Parser wandelt Dateien mit einem TXT-Format oder differierenden XML-Format aus fremden Bibliothekssystemen in das XML-Format des KOBV-Portals um.

Konvertierung über Konkordanzen

Wenn für den Import die Voraussetzung erfüllt ist und das KOBV-Portal-Format vorliegt, besteht die Hauptaufgabe des KMA-Parsers in der inhaltlichen Konvertierung bestimmter Metadaten-Elemente. Die Benennungen für die Fachgebiete, Ressourcetypen, Zugriffsarten und Sprachen, die in den lokalen Portalen individuell verwendet werden können, werden in die korrelierenden Benennungen des KOBV-Portals umgewandelt.

Im KOBV-Metadaten-Schema werden die ersten beiden Ebenen – zusätzlich wenige Ausnahmen aus der dritten Ebene – der DDC (Dewey Decimal Classification) in der deutschen Übersetzung¹² für die Klassifikation der Fachgebiete vorgeschrieben¹³. Da manche Fachgebiets-Benennungen der lokalen Bibliotheksportale nicht in die Klassifikation der DDC – und auch nicht in andere – zu konvertieren sind, weil sie kein Thema beschreiben, sondern vielleicht eine Zugehörigkeit zu einer Institution, können in diesen Fällen die Benennungen aus den Resourcebeschreibungen entfernt werden, ohne die übrigen Fachgebiets-Benennungen in irgendeiner Weise zu beeinträchtigen.

Bildung eines neuen Metadaten-Elementes

Im KOBV-Portal zeigen kleine „Icons“ in der Ressourcenliste an, ob die Ressourcen frei oder lizenziert sind, welcher Medientyp vorliegt und von welcher Bibliothek sie lizenziert sind, wenn es sich nicht um freie Ressourcen handelt. Diese Icons sind in den Ausgangsdaten nicht vorhanden und werden automatisch je nach Kontext der gelieferten Metadaten in die Resourcebeschreibungen für das KOBV-Portal eingetragen.

Synchronisation

Des Weiteren übernimmt der KMA-Parser die Synchronisation der Resourcebeschreibungen zwischen den lokalen Portalen und dem KOBV-Portal.

Ressourcen, die bis zu einem bestimmten Datum geändert wurden, werden erkannt, selektiert und in das KOBV-Portal übernommen. Gleiches geschieht mit den in den lokalen Portalen gelöschten Ressourcen: sie werden identifiziert und entsprechend aus dem KOBV-Portal entfernt.

Kontrolle und Auswahl

Im KMA-Parser verbergen sich auch Instrumente zur Kontrolle und Auswahl der Resourcebeschreibungen für das KOBV-Portal: Fehlerhaft oder unvollständig gelieferte Resourcebeschreibungen werden in Log-Dateien protokolliert. Dadurch wird eine Qualitätskontrolle für die KOBV-Zentrale und für die lokalen Bibliotheksportale gewährleistet. Da nicht alle Ressourcen aus den lokalen Bibliotheken in das Sammelprofil¹⁴ des KOBV-Portals passen, können im KMA-Parser nach frei benennbaren Kriterien Ressourcen ausgesondert werden und das sogar für jedes existierende Metadaten-Element.

¹¹ siehe <http://www.loc.gov/standards/marcxml/>

¹² Informationen zur deutschen Übersetzung von DDC siehe unter: http://www.ddb.de/professionell/ddc_info.htm

¹³ siehe Lavinia Hodoroaba, Andres Imhof: Das KOBV-Metadaten-Schema im KOBV-Portal. ZIB-Report 04-28, (August 2004) S.13 f. – Volltext unter: <ftp://ftp.zib.de/pub/zib-publications/reports/ZR-04-28.pdf>

¹⁴ Die KOBV-Zentrale hat eine Profilbeschreibung des KOBV-Portals veröffentlicht. Sie sieht darin einen Anfang und versteht diese als Aufruf an andere Portal-Betreiber, diesem Beispiel zu folgen. Ziel ist es, Sammelprofile und Zielgruppen abzugrenzen und einmal angelegte Metadaten gewinnbringend für alle Seiten auszutauschen. – Lavinia Hodoroaba, Andres Imhof, Markus Malo: Das Profil des KOBV-Portals. In: Bibliotheksdienst 39. Jg. (2005) H. 2. – Volltext unter: <ftp://ftp.zib.de/pub/zib-publications/reports/ZIB-04-54.pdf>

Außerdem werden freie Ressourcen aus den lokalen Portalen, die bereits im KOBV-Portal vorhanden sind, anhand der gleichen URL einander zugeordnet und für eine separat durchgeführte Überprüfung in eine andere Datei gespeichert.

Wechselseitiger Metadatenaustausch

Die genannten Aufgaben erfüllt der KMA-Parser nicht nur für den Import von Daten in das KOBV-Portal, dieser Vorgang funktioniert gleichermaßen in die entgegengesetzte Richtung zu den lokalen Bibliotheksportalen hin. Die einzelnen lokalen Bibliotheksportale können auf diese Weise die Ressourcebeschreibungen anderer Portale nutzen. Der KMA-Parser transformiert und konvertiert hierfür das KOBV-Metadaten-Schema in das jeweilige Schema der lokalen Bibliotheksportale.

Aufbau und Handhabung des KMA-Parsers

Der KMA-Parser wurde in Perl entwickelt und modular konzipiert, so dass einerseits einzelne Module auch in anderen KOBV-Projekten eingesetzt werden können (z. B. für die Synchronisation der einzelnen Konsortialportale) und andererseits funktionale Erweiterungen einfach implementiert werden können. Die Verarbeitung der XML-Dokumente erfolgt mit Hilfe des XML::DOM-Modules¹⁵ von CPAN¹⁶. Das XML::DOM-Modul baut das XML-Parser-Modul¹⁷ aus, das auf dem etablierten expat-XML-Parser¹⁸ von James Clark basiert. Weil das XML::Parser-Modul ein Non-Validating-Parser ist, findet die Validierung des XML-Dokumentes durch das XML::Validator::Schema-Modul statt, das ein XML-Dokument gegen ein Subset des W3C-XML-Schemas¹⁹ validiert.

Der KMA-Parser verfügt über vier Grundverzeichnissen, denen alle Dateien zugeordnet sind (siehe Abbildung 2):

- „conf“: enthält die Konfigurationsdatei und die Konkordanzen
- „data“: enthält jeweils einen Unterordnern pro Partner (lokales Bibliotheksportal u. a.) und für den KOBV, in denen die Ausgangsdaten abgelegt werden. Außerdem werden die Ausgangsdateien am Anfang des Prozesses im Unterordner „archive“ gespeichert.
- „ergebnisse“: Speicherung der neu erstellten Importdatei für das KOBV-Portal und der Log-Dateien im Unterordner „logs“
- „parser“: enthält den Code des KMA-Parsers

Wenn die Ressourcebeschreibungen eines lokalen Bibliotheksportals für das KOBV-Portal aufbereitet werden sollen, werden zunächst einmal im jeweiligen Portalsystem zwei vollständige Exporte der Daten des Partners und des KOBV-Portals ausgeführt. Es wird bei jedem Vorgang zum Abgleich immer auch ein aktueller, kompletter Datenabzug des KOBV-Portals benötigt, um feststellen zu können, ob im lokalen Portal eine Ressource entfernt wurde. Beide Export-Dateien werden anschließend in den entsprechenden Partner-Unterordner und in den KOBV-Unterordner des „data“-Verzeichnisses gespeichert. Von dort holt sich der KMA-Parser die Ausgangsdaten: die Export-Dateien des KOBV-Portals und des Partners. Dann bereitet der KMA-Parser die Partner-Daten nach den Konfigurationseinstellungen im Verzeichnis „conf“ auf, wo die Datei „grundeinstellungen.conf“ und sämtliche Konkordanzen der Metadata-Sharing-Partner für die Metadaten-Elemente „Fachgebiet“, „Ressourcetyp“, „Sprache“ und „Zugriffsart“ abgelegt sind. Abschließend speichert er die verbliebenen und veränderten Ressourcebeschreibungen in der neuen Import-Datei für das KOBV-Portal im Verzeichnis „ergebnisse“ ab. Gleichzeitig werden die Log-Dateien zum Umwandlungsprozess im Unterverzeichnis „logs“ abgelegt (siehe Abbildung 2).

¹⁵ <http://search.cpan.org/~timather/XML-DOM-1.43/lib/XML/DOM.pm>

¹⁶ <http://cpan.org>

¹⁷ <http://search.cpan.org/~msergeant/XML-Parser-2.34/Parser.pm>

¹⁸ <http://sourceforge.net/projects/expat/>

¹⁹ <http://www.w3.org/XML/Schema>

UNIX-Eingabe: `partner2kobv.pl ../data/kobv/kobv-datei.xml ../data/partner/partner-datei.xml`

Start: Enter

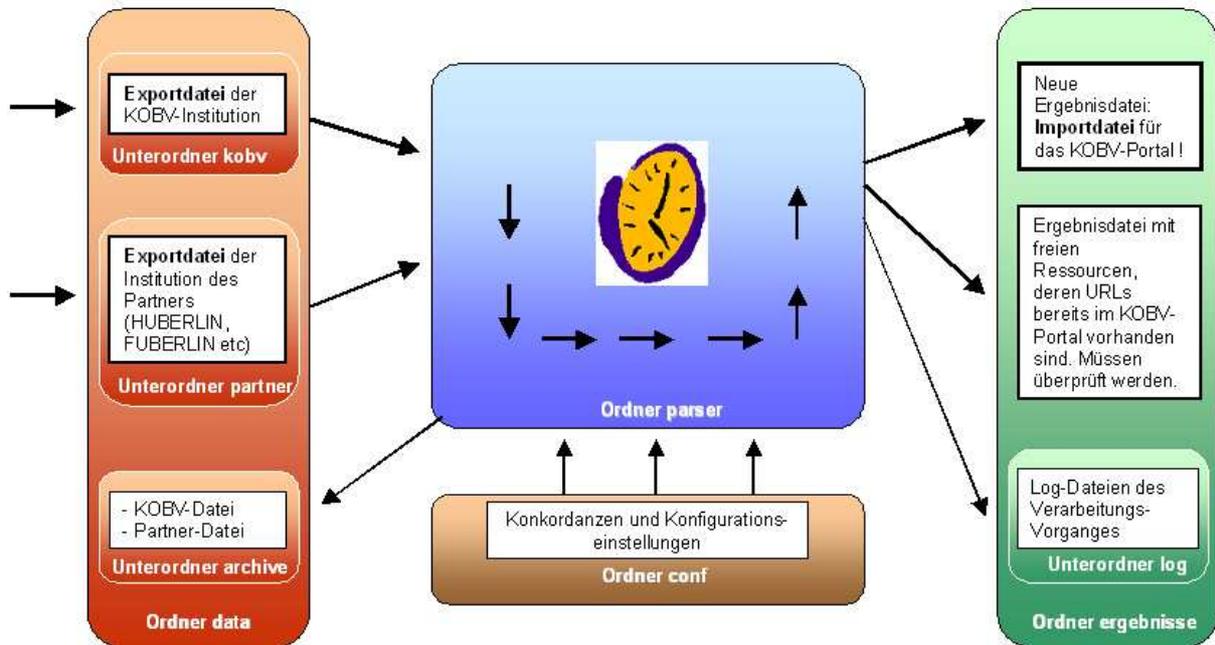


Abbildung 2: Handhabung und Vorgänge im KMA-Parser

Die Konfiguration

In der Datei „grundeinstellungen.conf“ des conf-Verzeichnisses werden alle Konfigurationen für den Umwandlungsprozess eingetragen. Diese Datei enthält für jeden Partner eine Sektion, die aus mehreren Zeilen besteht und für beide Richtungen des Umwandlungsprozesses gültig ist. Jede Zeile enthält einen Parameter, der im KMA-Parser während seines Bearbeitungsvorganges berücksichtigt wird (siehe Abbildung 3).

```

[euv]
kurzer_Name=EUV Frankfurt
vollstaend_Name=Europa-Universität Viadrina, Universitätsbibliothek
Adresse=Große Scharrnstr. 59, 15230 Frankfurt (Oder)
pruefe_Name=0#soll der Name geprueft werden?(0-nein, 1-ja)
interner_Partner=1# (0 = nein, 1 = Ja)
Praefix=EVR
Liste_cd_oder_www=#URL des lokal eingesetzten CD-ROM-Servers
Zeitraum=600#Zeitraum in Tagen
Konkordanz_Zugriffsart_in=zugriff_konkord_import.txt
Konkordanz_Zugriffsart_out=zugriff_konkord_export1.txt
Konkordanz_FIL_in=euv_fachgebiete_fil.txt
Konkordanz_FIL_out=euv_fachgebiete_fil.txt
Konkordanz_655_in=euv_typen_655.txt
Konkordanz_655_out=euv_typen_655.txt
Konkordanz_sprachen_in=sprachen_konkord_export2.txt
Konkordanz_sprachen_out=sprachen_konkord_export2.txt
#Erlaeuterung fuer Liste_Kriterien: #x(Feld_name,unterfeld,aktion,"Inhaltsfragment")x,
Liste_Kriterien=x(655, , ,a,record_nicht_einfuegen,"Bibliothekskatalog")x,x(245,1,
,a,record_loeschen,"BVG Fahrinfo online")x,x(245,1,
,a,record_nicht_einfuegen,"[Werkausgabe]")x
[ENDE] # das ist die obligatorisch letzte Zeile.

```

Abbildung 3: Datei „grundeinstellungen.conf“

Das Kürzel des Partners in rechteckigen Klammern leitet die Sektion der Bibliothek ein. Dann werden zunächst mit den ersten drei Zeilen der Kurzname, der vollständige Name und die Adresse der Bibliothek eingetragen. Die Adresse ist für eine lizenzierte Ressource einer Bibliothek notwendig, damit die NutzerInnen erfahren, an wen sie sich wenden müssen, wenn sie auf die Inhalte zugreifen möchten.

„pruefe_name“, „interner_partner“ und „praefix“ sind Parameter, die der KMA-Parser für die Zuordnung von internen Prozessen während des Umwandels benötigt. Z. B. zeigt der „interne Partner“, ob es sich um einen Konsortial-Partner handelt (Wert gleich 1).

Die „Liste_cd_oder_www“ gibt die URL eines CD-ROM-Servers der Bibliothek an, wenn ein solcher lokal eingesetzt wird. Der KMA-Parser trägt für die Kurzanzeige im KOBV-Portal automatisch Icons ein, die u. a. den Medientyp der Ressource (online oder CD-ROM) beschreibt. Er schließt auf eine CD-ROM, wenn keine URL in dem dafür vorgesehenen Metadatenfeld geliefert wird. Dies geschieht ebenso, wenn eine URL geliefert wird, die aber zum hier angegebenen CD-ROM-Server deklariert wurde. Alles andere ist eine Internet-Ressource.

Der „Zeitraum“ wird für die Bearbeitungsauswahl der neuen und veränderten Ressourcen im KMA-Parser herangezogen. Es werden nur die Datensätze im KMA-Parser selektiert und umgewandelt, die in den vergangenen angegebenen Tagen verändert worden sind.

Die Zeilen bezüglich der Konkordanzen geben den Namen der zu verwendenden Datei an. Meistens können die entsprechenden Konkordanz-Dateien für den Weg vom Partner zum KOBV-Portal und zurück vom KOBV-Portal zum Partner gleichermaßen genutzt werden. Es kann aber auch Ausnahmen geben, in denen sich die Konkordanzen unterscheiden. Für diese Fälle ist die Option eingerichtet worden, dass für die Umwandlungsrichtungen verschiedene Dateinamen angegeben werden können.

Und zu guter Letzt werden in „Liste_Kriterien“ die Angaben für die auszuschließenden Ressourcen festgehalten. Dabei gibt es zwei Aktionen: „record_loeschen“ und „record_nicht_einfuegen“. Die erste Aktion entfernt eine bereits im KOBV-Portal enthaltene Resourcebeschreibung. Die zweite Aktion

„record_nicht_einfuegen“ nimmt die entsprechende Ressource in der neuen Import-Datei gar nicht erst auf. Sie muss dementsprechend nicht importiert werden und vermindert somit die Menge der handzuhabenden Ressourcen.

Konkordanzen

Das Ziel des Einsatzes von Konkordanzen im KOBV-Portal ist, lokale individuelle Benennungen in Portalen zuzulassen, ohne dass dies einem gegenseitigen Austausch von Ressourcebeschreibungen im Wege steht. Das gilt zumindest für die Auswahl von Metadaten-Elementen, in denen festgelegte Begriffe, Klassifikationen oder Phrasen eingesetzt werden. Im KOBV-Portal sind dies die Fachgebiete, Ressourcetypen, Sprachen und die Zugriffsart. Die Konkordanzen beinhalten Konvertierungsregeln, die auf der einen Seite, die Benennungen des einen Austausch-Partners auflisten und auf der anderen Seite die Entsprechungen im KOBV-Portal nennen. Für die Konkordanzen werden TSV-Dateien²⁰ mit zwei Spalten eingesetzt. In der ersten Spalte steht der Wert der lokalen Bibliothek und in der zweiten Spalte der Wert des KOBV-Portals.

Meistens finden Konvertierungen 1:1 statt, wie z. B. bei den Sprachbenennungen. Bei den Fachgebieten-Klassifikationen muss aber auch ein Wert auf mehrere Werte abgebildet werden können, wenn beispielsweise das Fachgebiet „Anglistik“ in die Dewey Decimal Classifications (DDC) „420 Anglistik, Altenglisch“ (Sprache) und „820 Englische, altenglische Literatur“ gesplittet wird. Der KMA-Parser kann im Sinne der Konkordanz sowohl 1:1-Beziehungen, als auch 1:n-Beziehungen verarbeiten. Gleichzeitig prüft der KMA-Parser doppelte, gleichlautende Benennungen, wenn verschiedene Benennungen auf ein und denselben Wert konvertiert werden, und reduziert diese auf eine Benennung, damit im Portal keine Fachgebiete wiederholt werden (siehe Abbildung 4).

...	...
Anglistik	420 Anglistik, Altenglisch
Anglistik	820 Englische, altenglische Literatur
Biologie	570 Biowissenschaften; Biologie
Kunstgeschichte	700 Künste
Land und Forstwirtschaft	630 Landwirtschaft
Gartenbau	630 Landwirtschaft
Institutsbibliothek1	feld_loeschen
Institutsbibliothek2	feld_loeschen
...	...

Abbildung 4: Konkordanz für die Fachgebieten-Benennungen: euv_fachgebiete_fil.txt

Obwohl eine Konkordanz in beide Richtungen eingesetzt werden kann, ist im KMA-Parser die Option vorhanden, für den Hin- und Rückweg zwei verschiedene Konkordanzen zu verwenden. Es kann sein, dass eine Bibliothek aus internen Gründen mehrere unterschiedliche Benennungen für ein und dasselbe Fachgebiet liefert. Für den Rückweg muss jedoch eine klare Zuordnung der DDC-Klassifikation auf einen dieser Bibliotheksbenennungen vorgenommen werden. Das kann über die getrennte Angabe einer neuen Konkordanz geschehen.

Es bleibt noch eine Funktion der Konkordanzen zu erwähnen, die für Fachgebiete und Ressourcetypen eingerichtet wurde. Da lokal individuelle Klassifikationen eingesetzt werden, können Begriffe vorkommen, die nicht in die Benennungen des KOBV-Metadaten-Schemas übersetzt werden können. Das ist der Fall, wenn beispielsweise eine Bibliothek alle Ressourcen jeweils namentlich einzelnen Teilbibliotheken zuordnet. Diese Benennungen erfüllen ihren Zweck im lokalen Portal, stören aber in einem regionalen Portal, da man eine Teilbibliothek nicht in die vom KOBV-Portal eingesetzte deutsche Übersetzung der DDC übertragen kann. Für solche Fälle wurde die Funktion „feld_loeschen“ implementiert. Wenn hinter einer lokalen Benennung diese Funktion steht, wird der Begriff aus dem Datensatz entfernt, ohne die anderen Klassifikationen in Mitleidenschaft zu ziehen

²⁰ tsv = tabulator separated value

(siehe Abbildung 4). Für den Fall, dass ursprünglich nur eine Benennung im Datensatz vorhanden ist, wird allerdings die Überprüfung der obligatorischen Metadatenfelder hier anschließend einen Fehler feststellen, da sowohl das Fachgebiet als auch der Ressourcetyp ein obligatorisches Metadaten-Element ist, und den Datensatz entsprechend aussondern.

Erfahrungen und Ergebnisse

Da zum gegenwärtigen Zeitpunkt nur wenige Bibliotheken über strukturierte Metadaten zu ihren lizenzierten Ressourcen verfügen und die Erfahrungen in diesem Bereich noch sehr gering sind, waren Fortschritte im Projekt „KOBV-Portal II - Metadata-Sharing“ zunächst nur sehr zäh zu erreichen. Zumindest konnte die KOBV-Zentrale auf die Ressourcebeschreibungen zweier Partnerbibliotheken zurückgreifen, die zeitgleich zum KOBV-Portal lokale Bibliotheksportale mit MetaLib aufbauten. In den Bibliotheken selbst konnte nur wenig Zeit aufgebracht werden, auf übergeordnete Zusammenhänge im Sinne von Metadata-Sharing einzugehen, da die Entwicklung der lokalen Bibliotheksportale alle Aufmerksamkeit und Kräfte auf sich zogen. Beide Portale wurden somit individuell nach lokalen Gesichtspunkten konzipiert. Die Konsequenz für das KOBV-Portal lag darin, dass der KMA-Parser an die individuellen Vorgehensweisen in den lokalen Bibliotheksportalen angepasst werden musste. Das erschwerte zwar insgesamt die Konzeptionierung des KMA-Parsers und verlängerte deutlich die Entwicklungszeit, dadurch wurde er aber auch erheblich leistungsstärker und ermöglicht jetzt – bis zu einem bestimmten Rahmen – die Unabhängigkeit des KOBV-Portals von den lokalen Bibliotheksportalen.

Die KOBV-Zentrale ist nun in der Lage, anhand des Zusammenspiels unterschiedlichster Datensätze grundlegende Prinzipien im Umgang mit heterogenen Datensätzen als Ergebnisse des Projektes vorzustellen.

Unterstützung von lokalen Benennungen und Standards gleichermaßen

Über den KMA-Parser können dezidiert Metadatenfelder, die mit eindeutig festgelegten Benennungen nach lokalen Ordnungen oder nationalen bzw. internationalen Standards beschrieben werden, jederzeit über entsprechende Konkordanzen in jeden nur erdenklichen Standard umgewandelt werden. Bibliotheken und anderen Betreibern von Portalen weltweit ist es auf diese Weise freigestellt, welche Formate und Klassifikationen sie für die einzelnen Metadaten-Elemente einsetzen. Dies ist eine wesentliche Anforderung und entspricht der konkreten Praxis, weil natürlich jeder Portalanbieter Inhalte und Darstellungsweise nach eigenen Ansprüchen selbst bestimmen möchte. Das klassische Beispiel hierfür sind die Fachgebiete, für die die unterschiedlichsten Klassifikationen angesetzt werden. Aber auch Ressourcetypen, für die es noch keinen wirklich geeigneten Standard gibt, werden individuell benannt. In einem Netzwerk von Portalen können mit dem KMA-Parser einmal erstellte Ressourcebeschreibungen ohne Hindernisse an alle Teilnehmer ausgeliefert werden und zwar im jeweils korrekten Metadatenformat. Allerdings ist bei diesem Vorgang, der Umwandlung von Informationen, generell zu berücksichtigen, dass häufig ein Verlust von Informationen eintritt.

Qualitätskontrolle der Ressourcebeschreibungen

Zu Beginn des Projektes fehlten zunächst die Informationen zu den individuellen Fachgebietsbenennungen in den Bibliotheksportalen. Die jeweiligen Konkordanzen mussten darum in der KOBV-Zentrale selbstständig generiert werden. Wie sollte man aber an die kompletten Benennungen der bis zu 650 einzelnen Datensätze pro Bibliothek herankommen? Für diesen Zweck wurde im KMA-Parser die Protokollierung aller erfolgreichen und nicht erfolgreichen Umwandlungsvorgänge in Log-Dateien eingeführt. Anhand dieser Log-Dateien sieht man beispielsweise, ob alle obligatorischen Metadaten geliefert wurden und für welchen Datensatz die Fachgebiete, Ressourcetypen etc. unbekannt oder die Schreibweisen fehlerhaft sind. Um eine neue Konkordanz aufzustellen, werden jetzt die Datensätze durch den KMA-Parser umgewandelt, ohne dass eine Konkordanz vorhanden ist. Anschließend werden die Fachgebiete, Ressourcetypen usw. in der Log-Datei als Fehlermeldungen genannt, die von der jeweiligen Bibliothek verwendet werden. Die Log-Dateien dienen also gleichzeitig als Grundlage für die Aufstellung einer neuen oder korrigierten Konkordanz und zur Korrekturhilfe in den lokalen Bibliotheksportalen. Auf diese Weise liefert der KMA-Parser einen bedeutenden Beitrag zur Qualitätsprüfung der Ressourcebeschreibungen für alle Partner des Metadata-Sharings.

Kontrolle über Ressourcenauswahl nach selbst bestimmten Kriterien

Anhand der Zusammenführung lokaler Ressourcen für ein universelles, regionales Portal wurde sehr bald deutlich, dass es unterschiedliche Ziele von Portalen gibt. Sie unterscheiden sich in der Zielgruppe und im Sammelpprofil, also der thematischen Auswahl an Ressourcen und den damit zusammenhängenden Ressourcetypen. Diese Ziele können in einer Profilbeschreibung zusammengefasst werden, wie sie beispielsweise die KOBV-Zentrale für das KOBV-Portal veröffentlicht hat²¹.

Ein lokales Bibliotheksportal nimmt andere Ressourcen auf, als sie ein regionales Portal für relevant erachtet, eine Virtuelle Fachbibliothek führt Ressourcen mit einem tief erschlossenen Themenausschnitt, während ein universell ausgerichtetes Portal nur die wichtigsten Ressourcen aller Fachgebiete anbieten möchte. Die Integration lokaler Datensätze anderer Portale kann also nur geschehen, wenn auch eine einfach zu konfigurierende Kontrollmöglichkeit über die Auswahl nach selbst bestimmbareren Kriterien zur Verfügung steht. Schließlich sollen die Datensätze in großem Umfang über automatisierte Prozesse ausgetauscht werden, ohne jeden einzelnen Datensatz anpacken zu müssen. Diese Kontroll- und Auswahlfunktion wurde im KMA-Parser in Form der Kriterienauswahl zum Ausschluss von Ressourcen integriert.

Empfehlung zur Berücksichtigung von Standards bei allen Partnern

Es wurde bereits konstatiert, dass durch die technischen Möglichkeiten des KMA-Parsers das KOBV-Portal in einem bestimmten Rahmen von den lokalen Bibliotheksportalen unabhängig ist. Die einmal in den Bibliotheken angelegten Datensätze können ohne großen Arbeitsaufwand automatisch mit nur wenig Konfigurationsaufwand in das KOBV-Portal übernommen werden. Dennoch gibt es einzelne Fälle, wo auch der KMA-Parser an seine natürlichen Grenzen stößt.

Dies ist der Fall, wenn beispielsweise eine Bibliothek eine Ressource dupliziert, um sie gleichzeitig zwei unterschiedlichen Ressourcetypen zuzuordnen. Das ist notwendig, wenn für einen Datensatz – wie es in MetaLib der Fall ist – nur die Vergabe von einem Ressourcetyp zulässig ist, eine Bibliothek aber zwei oder mehrere vergeben möchte. Der KMA-Parser kann diese Ressourcen nicht differenzieren, da sich außer der Benennung des Ressourcetyps keine inhaltliche Unterscheidung findet. Demzufolge erscheinen beide Ressourcebeschreibungen parallel im KOBV-Portal.

Der KMA-Parser erkennt anhand der URL des eingetragenen CD-ROM-Servers einer Bibliothek, dass es sich um eine CD-ROM handelt. Manch eine Bibliothek steuert jedoch über die gleiche URL auch lizenzierte Zugänge zu Online-Datenbanken an. Diese Ressourcen erscheinen im KOBV-Portal als CD-ROM, nicht als Internet-Ressource, was inhaltlich falsch ist und korrigiert werden muss.

CD-ROMs, die einer gemeinsamen Kategorie angehören, wie z. B. die Werkausgabe eines Autors, wären einfacher, mit weniger Konfigurationsaufwand für ein regionales Portal auszuschließen, wenn in den Bibliotheken selbst standardisierte Benennungen und Kategorisierungen eingehalten würden, anhand derer alle zutreffenden CD-ROMs erkannt werden könnten. Dies ist nicht immer der Fall, darum müssen zum Teil einzelne CD-ROM-Ressourcen allein anhand des Titels selektiert werden. Dies erhöht den Konfigurationsaufwand unnötig.

Fazit

Der KMA-Parser bietet aufgrund seiner vielfältigen Funktionen eine bedeutende Unabhängigkeit von den gelieferten strukturierten Metadaten. Die gelieferten Informationen aus der Region werden für die Darstellung im KOBV-Portal optimiert. Aber auch in die Gegenrichtung bietet der KMA-Parser die Auslieferung von Metadaten an andere Portale, die zu einer Vernetzung der Portale in Bezug auf Metadata-Sharing nicht nur innerhalb des KOBV führen kann. Der Austausch strukturierter Informationen durch den KMA-Parser ist für alle standardisierten Formate möglich. Er verfügt damit über eine universelle Offenheit für das Metadata-Sharing.

Der KMA-Parser stößt auch an seine Grenzen. In den genannten Beispielen ist die KOBV-Zentrale jeweils auf die Kooperation der lokalen Bibliotheksportale angewiesen. In einigen Metadatenfeldern

²¹ Lavinia Hodoroaba, Andres Imhof, Markus Malo: Das Profil des KOBV-Portals. In: Bibliotheksdienst 39. Jg. (2005) H. 2. – Volltext unter: <ftp://ftp.zib.de/pub/zib-publications/reports/ZIB-04-54.pdf>

sind Freiheiten zulässig und sinnvoll, in anderen sind dennoch Absprachen zwischen den einzelnen Partner unbedingt notwendig.

Diese Erfahrung weist auf eine grundlegende Perspektive in der Zusammenarbeit von Informationsanbietern und - Providern hin: Werkzeuge wie der KMA-Parser schaffen eine gewisse Unabhängigkeit im Ressourcenaustausch mehrerer Partner, dennoch funktioniert dies nur für einen Teilbereich der Metadaten. Andere Inhalte von Metadaten-Elementen sollten untereinander abgestimmt werden. Die Metadaten-Austausch-Formate müssen sich an nationalen oder internationalen Standards (beispielsweise MABxml²², MARCXML²³ oder BiblioML²⁴) orientieren, damit Transformationen im Vorfeld unnötig werden. Der Königsweg für ein weltweites Metadata-Sharing könnte sein, einerseits mittels Konkordanzen lokale Benennungen zu bedienen und andererseits große Bestandteile der Metadaten, Schemata und Formate auf Standards hinzuführen, um so die Informationsflüsse auf einheitlichen Wegen für eine übergreifende Verwendung zu kanalisieren.

²² siehe <http://www.ddb.de/professionell/mabxml.htm>

²³ siehe <http://www.loc.gov/standards/marcxml/>

²⁴ siehe <http://xml.coverpages.org/biblioML.html>