

Katalogisierung kooperativ gemacht

von

Barbara Block, Josef Labner, Beate Rusch

Stand: 15.01.2007

Haben Sie schon einmal versucht, Ihren fachfremden Bekannten oder Familienangehörigen plausibel zu erklären, warum in Deutschland (die Österreicher mögen diese Verkürzung einen Moment verzeihen und sich mitgenannt fühlen) eine Titelaufnahme zeitgleich mehrfach erstellt wird? Mal unter uns: Wir sind mit unseren ausführlichen Erklärungen regelmäßig gescheitert. Spätestens wenn wir mit unserem Vortrag bei unterschiedlichen Katalogisierungskonventionen und technischen Hindernissen angelangt sind, ist uns unser Gegenüber eingnickt. Unser fachfremdes Gegenüber, versteht sich.

Aber je stärker die Katalogisierungsabteilungen in den Bibliotheken schrumpfen und die „pragmatische“ der „perfekten“ Titelaufnahme den Rang abläuft, desto mehr Fragen kommen auch von BibliothekarInnen: Warum lassen sich die Aufnahmen nicht aus anderen Verbänden übernehmen? Warum nicht die verknüpften Normdaten? Warum existieren nach wie vor unterschiedliche Erschließungspraktiken?

Die Arbeitsgemeinschaft der Verbundsysteme hat sich 2005 dieser Thematik angenommen und ein Projekt mit dem Namen „Kooperative Neukatalogisierung“ ins Leben gerufen.¹ Es sollten Maßnahmen zur Verbesserung der Datenübernahme zwischen den Verbänden für die Formal- und Sacherschließung erarbeitet werden mit dem Ziel, die Eigenkatalogisierungsquoten zu senken sowie Anreicherungen für bestehende Datensätze (Sacherschließung, Links zu Rezensionen und Inhaltsverzeichnissen usw.) zu ermöglichen. In einem zweiten Paket sollte die Angleichung bzw. Vereinheitlichung von differierenden Regelwerks- und Formatanwendungen erreicht und ein gemeinsamer Qualitätsstandard in den Verbänden vereinbart werden.

An diesem ambitionierten Unterfangen sind die Mitglieder der Arbeitsgemeinschaft der Verbundsysteme (die Verbände BVB, GBV, hbz, HeBIS, KOBV, OBV, SWB und ZDB und die Deutsche Nationalbibliothek (DNB)) beteiligt. Die Deutsche Nationalbibliothek hat in diesem Projekt eine besondere Ausgangslage. Während die Nutzung der Daten der Deutschen Nationalbibliothek in den Verbänden die Regel ist, betritt die Deutsche Nationalbibliothek Neuland, wenn sie im Rahmen dieses Projektes prüft, in welchem Umfang Verbunddaten für die eigene Arbeit genutzt werden können.

¹ Es wurde eine Projektgruppe eingerichtet, in der Experten aus den Bereichen Regelwerke, Datenformate, Z39.50/Online-Kommunikation und IT-Spezialisten aus den Verbänden und der DNB zusammenarbeiten.

Mehrere Wege führen zum Ziel

Die Nutzung von Fremddaten wird in den Verbundsystemen unterschiedlich organisiert. Hier führen mehrere Wege zum Ziel. In Aleph-Verbänden (BVB, hbz, KOBV, OBV) gehört neben der Recherche in verbundeigenen Fremddatenpools die Datenübernahme über Z39.50 zu den Standardverfahren. Hier stehen ausgereifte Z39.50-Recherche-Schnittstellen zur Verfügung und die Suche in deutschsprachigen Verbunddatenbanken wird in der Regel seit längerem angeboten.

In Pica-Verbänden (GBV, HeBIS, SWB) wird dagegen eine integrierte Lösung für die Fremddatennutzung bevorzugt: Fremddaten werden zusammen mit den Verbunddaten im Pica-Zentralsystem bereitgehalten.

Diesen unterschiedlichen verbundspezifischen Gegebenheiten wurde im Projekt Rechnung getragen. So entschied man sich dafür, sich zum einen die Daten gegenseitig per Lieferungen (ftp bzw. OAI) bereitzustellen und zum anderen für die Online-Übernahme auch die Z39.50-Schnittstellen zu optimieren.

Einheitliche Suchumgebung auch in fremden Datenbanken

Die deutschsprachigen Verbände wie auch die Deutsche Nationalbibliothek haben ihre Datenbanken für die Recherche über Z39.50 geöffnet und damit die technischen Voraussetzungen geschaffen, ihre Kataloge in fremde Suchumgebungen zu integrieren. Allerdings ist mit der Realisierung dieser Schnittstelle alleine eine Harmonisierung von Sucheinstiegen noch nicht erreicht. Dazu bedarf es weiterer Vereinbarungen, einer Auslegung der Schnittstelle gewissermaßen, die in so genannten Z39.50-Profilen erfolgt. Ein solches gemeinsames Z39.50-Profil legen die deutschsprachigen Verbände und die DNB nun im Rahmen des Projektes erstmals vor. Darin enthalten ist eine einheitliche Definition von Sucheinstiegen und Indices. Das Profil orientiert sich an dem international registrierten Bath Profile² und ist ausgerichtet auf die Bedürfnisse von KatalogisiererInnen, die Fremddaten recherchieren und übernehmen möchten, sowie von Portalbetreibern, die für ihre Benutzer eine verteilte Suche über mehrere Kataloge anbieten.

Das Profil umfasst die folgenden Sucheinstiege: Autor (Personen und Körperschaften), Titel, Schlagwort, Standardnummern, Jahreszahl und ergänzend zum Bath Profile getrennte Sucheinstiege jeweils für persönliche Verfasser und körperschaftliche Verfasser. Das Profil, das in der „standardisierten Sprache“ des Z39.50-Protokolls abgefasst ist, wird in Kürze nachzulesen sein auf den Web-Seiten der AG der Verbundsysteme.³

² Das international registrierte Bath Profile lässt sich nachlesen unter: <http://www.collectionscanada.ca/bath/ap-bath-e.htm>.

³ S. unter <http://www.ag-verbund.de>

Ergänzend zu dem Z39.50-Targetprofil legen die deutschsprachigen Verbände und die Deutsche Nationalbibliothek erstmals eine einheitliche Definition für die Indexierung der im Targetprofil beschriebenen Sucheinstiege vor.⁴ Diese Indexierungsvereinbarung regelt die Inhalte, die in den jeweiligen Stichwort- bzw. Phrasenindex einfließen sollen. Dabei wird ein Kernset als Mindeststandard formuliert. Darüber hinausgehende Erweiterungen sind selbstverständlich möglich.

Folgende Felder sollen für die jeweiligen Indices berücksichtigt werden:

- Für den Personen-Index die MAB-Felder 100, 104, 108 ..., 196 und die jeweiligen Verweisungsformen der Personennamen
- Für den Körperschaften-Index die MAB-Felder 200, 204, 208 ..., 296 und die jeweiligen Verweisungsformen der Körperschaftsnamen
- Für den gemeinsamen Personen- und Körperschaften-Index die MAB-Felder 100, 104, 108 ..., 196, 200, 204, 208 ..., 296 und die jeweiligen Verweisungsformen der Personen- und Körperschaftsnamen
- Für den Titel-Index die MAB-Felder 089, 304, 310, 331, 335, 341, 345, 349, 353, 360, 370, 670, 675
- Für den Standardnummern-Index die MAB-Felder 025a, 025l, 025z, 026Blank, 376b, 540, 541, 542, 543, 552, 556, 634, 635
- Für den Index „Erscheinungsjahr“ die MAB-Felder 425, 619
- Für den Index „Alle Felder“ die folgenden MAB-Felder: alle Felder der anderen o.g. Indices und zusätzlich Ortsangaben (MAB-Felder 410, 415, 611, 614), Verlagsangaben (MAB-Felder 412, 417, 613, 616) sowie Sacherschließung (hierfür wurden noch keine MAB-Felder festgelegt).

Die Realisierung des gemeinsamen Kernsets erfordert von den Projektpartnern eine (unterschiedlich umfangreiche) Re-Indexierung der Datenbanken. Als Zeitpunkt für die Realisierung des gemeinsamen Kernsets und des Targetprofils durch alle Projektpartner wurde der 01. Februar 2007 festgelegt. Ausgenommen ist dabei der gemeinsame Personen- und Körperschaften-Index, der erst zu einem späteren Termin in allen Systemen realisiert wird. Mit dem vorliegenden Profil und der entsprechenden Harmonisierung eines Kernsets von Sucheinstiegen wird die Recherche in fremden Datenbanken transparenter und damit zuverlässiger und komfortabler als bisher.

Einheitliche Datenlieferungen aus fremden Datenbanken

⁴ Für den Bereich der Sacherschließung wurden noch keine Festlegungen getroffen, da der Arbeitsschwerpunkt zunächst bei den Indices lag, die für die Formalkatalogisierung benötigt werden. Festlegungen für den Bereich der Sacherschließung sollen zu einem späteren Zeitpunkt ergänzt werden.

Damit der Zugriff auf alle Neukatalogisate tatsächlich auch allen KatalogisiererInnen in allen Verbänden sowie in der Deutschen Nationalbibliothek zu einem möglichst frühen Zeitpunkt garantiert wird, haben die Projektteilnehmer auch formell vereinbart, sich ihre Daten gegenseitig zur Nachnutzung über Z39.50 und darüber hinaus in Form von Datenlieferungen zur Verfügung zu stellen. In diesem zweigleisigen Verfahren spiegeln sich die unterschiedlichen Ansätze der Fremddatennutzung.

Ab Mai 2007 werden von allen Projektpartnern (mit Ausnahme der DNB) im Routinebetrieb wöchentlich über ftp, in der zweiten Stufe über OAI Differenzabzüge mit neuen und korrigierten Datensätzen geliefert. Dabei sollen insbesondere auch Bestellkatalogisate zur Verfügung gestellt werden. Diese sind gesondert zu kennzeichnen. Handelt es sich um die Aufnahme eines Bandes, werden über ftp vollständige Austauschseinheiten bereit gestellt (d.h. es werden die Überordnungen zu mehrbändigen begrenzten Werken mitgeliefert), über OAI ist das vermutlich nicht zu realisieren.

Die Datendienste der DNB bleiben zunächst im gewohnten Lieferumfang und -rhythmus abrufbar, OAI als „schneller Lieferweg“ wird von der DNB im Laufe des Jahres 2007 angeboten werden.

Nicht ausgetauscht werden Änderungen, die aufgrund von maschinellen Korrekturen entstehen und sehr viele Daten betreffen können, sowie Umlenkungen und Löschungen. Es ist dem empfangenden Verbund bzw. der DNB vorbehalten, die Daten weiter zu selektieren und nicht erwünschte Daten zu separieren.

Für das Datenformat wurde eine Reihe von Festlegungen getroffen, die sowohl für den Export über Z39.50 als auch für die Datenlieferungen gelten. Ausgetauscht wird zunächst im Format MAB2, später in MARC 21. Für die Datenübernahme von grundlegender Bedeutung ist die Festlegung, dass ausschließlich überregionale Normdatenidentifikationsnummern (sofern vorhanden) ausgegeben werden sollen, da regionale bzw. lokale Identifikationsnummern nur für die eigene Datenbank von Bedeutung sind.⁵ Aus diesem Grund wird empfohlen, Serien in der ZDB zu erfassen. Für eine vollständige Nutzung aller von der DNB angelegten Aufnahmen ist die schnelle Integration der Serienaufnahmen der DNB in die ZDB ein dringendes Desiderat.

⁵ Bei der Nachnutzung von Formal- und Sacherschließungsdaten ist es sehr wichtig, dass auch die Verknüpfungen zu Normdateien nachgenutzt werden können. Ziel muss es sein, möglichst viele Sätze in den überregionalen Normdateien nachzuweisen. Die Online-Schnittstelle soll in Zukunft die Erfassung in den überregionalen Normdateien sicherstellen. Daneben ist zu berücksichtigen, dass es in einigen Verbänden noch zahlreiche Personen- und Namenssätze gibt, die nicht in der überregionalen PND vorhanden sind. Hier besteht akuter Handlungsbedarf!

Die ursprüngliche Herkunft des Datensatzes soll am MAB-Feld 026 (Indikator Blank) erkennbar sein. Dieses Feld wird mit der Identifikationsnummer des ersterfassenden Verbundes bzw. der DNB besetzt. Durch die Zuordnung einer eindeutigen Identifikationsnummer zu einem Datensatz können nachträgliche Korrekturen und Anreicherungen aus allen anderen Verbänden übernommen werden. Für Daten, die auf Fremddaten von der BNB, Casalini, LoC usw. basieren, wird in MAB-Feld 025 die Identifikationsnummer des Fremddatenlieferanten mit dem entsprechenden Indikator angegeben.

Im Rahmen des Projektes wurde auch das Lieferverfahren der DNB an die Verbände diskutiert. Dabei wurde deutlich, dass alle Verbände eine schnellere Lieferung der Titeldaten wünschen. Diesem Wunsch will die DNB beim Lieferverfahren über OAI entgegenkommen. Im Laufe des Jahres 2007 soll es soweit sein.

Ein genialer Identifier

Die bislang beschriebenen Maßnahmen – Harmonisierung der Suche über Z39.50 und regelmäßige Differenzabzüge – haben die Bestellkatalogisierung im Fokus. Hier muss es das Ziel sein, möglichst schnell einen geeigneten Datensatz zu finden. Doch welches Verfahren lässt sich in den bestehenden dezentralen Strukturen finden, in dem alle Partner auch von (nachträglichen) Korrekturen und den Anreicherungen (Sacherschließung, Links zu Rezensionen und Inhaltsverzeichnissen usw.) anderer profitieren? Basierend auf bereits vorhandenen Replikationsmechanismen in den Aleph-Verbänden entwickelte die Arbeitsgruppe einen einfachen Mechanismus für eine kooperative Katalogisierung, durch den auch Korrekturen und Anreicherungen allen Projektpartnern zu gute kommen. Die derzeit zur Verfügung stehenden Funktionen mussten dabei nur geringfügig erweitert werden.

Der Charme liegt dabei in dem minimalen Implementierungsaufwand bei einem gleichzeitig hohen Nutzeffekt. Allerdings ist das vorgestellte Verfahren kein „100%-Verfahren“; nicht alle denkbaren Fälle werden abgedeckt. Es basiert vielmehr auf der Prämisse, dass schon für die Bestellkatalogisierung verbundübergreifend Fremddaten in großem Stil genutzt werden. „Absichtliche“ Doppelerfassungen des gleichen Titels, bei denen ein Eigenkatalogisat erstellt wird, statt ein Erstkatalogisat aus einem anderen Verbund zu übernehmen, werden von dem Modell nicht abgedeckt.

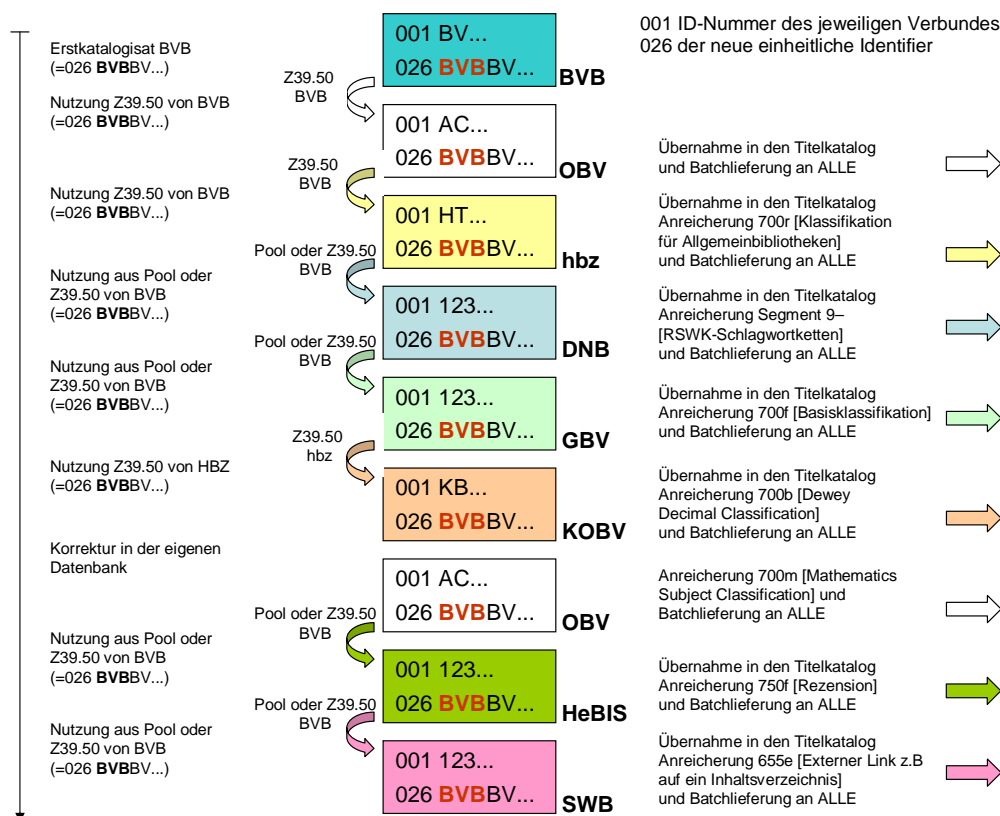
Das „Verfahren 026“

Bereits jetzt ist es möglich (und durchaus üblich) bei der Nutzung eines Fremddatensatzes die Identifikationsnummer der Quelle (MAB2-Feld 001) in eine andere Kategorie des Datensatzes (meist MAB2-Feld 026) zu übernehmen. Diese Information kann dann - wie bei

den im OBV bereits implementierten automatischen Nachführungen der DNB-Schlagwortketten - für ein Updateverfahren genutzt werden. Dieses Modell funktioniert allerdings nur dann, wenn die Anreicherung/Verbesserung des Katalogisats durch jenen Verbund erfolgt, von dem der Datensatz auch übernommen wurde. Das ist jedoch leider nicht die Realität. So mag zwar Verbund 1 das erste Erwerbungs-katalogisat anfertigen, aber Verbund 2 hat beispielsweise die schnellste Sacherschließung.

Ziel sollte jedoch sein, die Anreicherungen aller Verbünde nachnutzen zu können. Dazu wiederum benötigt man eine gemeinsame Identifikationsnummer, die von allen Beteiligten mitgeführt wird. Auf der Basis dieser gemeinsamen Identifikationsnummer können dann einfach Updates eingespielt und Anreicherungen in den einzelnen Verbänden nachgeführt werden.

Verfahren mit einer einheitlichen Identifikationsnummer (Kategorie 026 Indikator Blank):



Der in dem Schema abgebildete Datenfluss besteht nicht aus einer Kette, die nicht unterbrochen werden darf, sondern basiert auf der – über alle Verbünde hinweg – einheitlichen Identifikationsnummer 026 (Indikator Blank), die der erstkatalogisierende Verbund vergibt. Dabei wird davon ausgegangen, dass alle Daten allen zur Verfügung

stehen, entweder über eine Z39.50-Recherche oder über gegenseitige Datenlieferungen, die in verbundinterne (Fremd-)Datenpools eingespielt werden.⁶

Im obigen Beispiel wird das Erstkatalogisat vom BVB zunächst über Z39.50 durch den OBV genutzt. Das hbz nutzt den Datensatz ebenfalls über Z39.50 vom BVB und ergänzt die Klassifikation für Allgemeinbibliotheken. Die DNB nutzt den Datensatz (entweder über Z39.50 oder aus ihrem eigenen Fremddatenpool) und ergänzt um die Schlagwortketten. Der GBV hat den Datensatz in seinen Datenpool (versehen mit 026 Blank) eingespielt und ergänzt den BVB-Datensatz mit einer Basisklassifikation.

Der KOBV wiederum wird über Z39.50 beim hbz fündig und ergänzt die DDC.

Der OBV ergänzt den Datensatz noch einmal um die Mathematics Subject Classification.

HeBIS nutzt den Datensatz (aus dem eigenen Pool) und ergänzt um eine Rezension. Dann liefert der SWB zu diesem Datensatz noch einen externen Link auf das Inhaltsverzeichnis der Publikation.

Alle Anreicherungen kommen allen Verbänden zu Gute - Dank des Identifiers 026!

Konkrete Festlegungen für das „Verfahren 026“

Folgende konkrete Festlegungen wurden getroffen: Das Feld MAB 026 (Indikator Blank) ist obligatorisch beim Datenausch und nicht wiederholbar.

Es wird mit der Identifikationsnummer des ersterfassenden Verbundes bzw. der DNB besetzt. Entweder wird in 026 die eigene Identifikationsnummer geliefert oder – falls der Datensatz von einem anderen Verbund übernommen worden war – die Identifikationsnummer des ursprünglich erfassenden Verbundes bzw. der DNB.

Der Identifier setzt sich aus einem 3-stelligen Präfix und den alphanumerischen Identifikationsnummern der Verbände bzw. der DNB zusammen. Er enthält kein Trennzeichen. Als Präfixe sind erlaubt: BSZ, BVB, DNB, GBV, HBZ, HEB, KBV, OBV, ZDB. Die Identifikationsnummern von anderen Fremddatenlieferanten werden in MAB 025, 568 und 574 mit den entsprechenden Indikatoren geliefert.

Auch das MAB-Feld 070 (Indikator a) wird für die liefernde Institution obligatorisch und mit einem Code besetzt, der den oben genannten Präfixen entspricht.

Die MAB-Felder 001 bzw. 010 werden – wie gewohnt - mit der eigenen Identifikationsnummer besetzt.

⁶ Die ZDB beteiligt sich ebenso wie die anderen Verbände und die DNB an dem Verfahren 026. Für die Zeitschriften (bzw. die fortlaufenden Sammelwerke insgesamt) ist der Datenfluss aufgrund der gemeinsamen Katalogisierung in eine zentrale Normdatei jedoch anders organisiert. Daher wurde die ZDB in das obige Schema nicht mit aufgenommen.

Die Projektpartner verpflichten sich, die Kategorie 026 (Indikator Blank) ab dem 1. März 2007 zu liefern. Dies bezieht sich sowohl auf Erstkatalogisate, als auch auf alle Titelkorrekturen. Titelsätze mit Ersterfassungsdatum vor diesem Stichtag sind hinsichtlich des erstkatalogisierenden Verbundes nicht mehr eindeutig zuzuordnen, richten jedoch bei den Update-Verfahren auch keinen „Schaden“ an, sondern laufen schlicht ins Leere.

Zusammenfassung und Ausblick

Mit dem von Pragmatismus getragenen Projekt Kooperative Neukatalogisierung ist den deutschsprachigen Verbänden und der Deutschen Nationalbibliothek ein virtueller Schulterschluss gelungen, der sich für die KatalogisiererInnen in allen Verbänden und in der DNB lohnen wird. Ein Bündel von Maßnahmen wurde erarbeitet, das nun bis zum 1. Mai Schritt für Schritt umgesetzt wird. Die verbesserte, da vereinheitlichte Recherche über Z39.50 hat bereits kurzfristig Aussicht auf Erfolg und Akzeptanz. Das gleiche gilt für die Datenlieferungen, die sich als vergrößerter Fremddatenpool bei den PICA-Verbänden wiederfinden werden. Im Mai soll der Routinebetrieb erreicht sein und sich die Situation der Datenübernahmen in allen Verbänden spürbar verbessert haben.

Das Anreicherungsverfahren über MAB 026 (den genialen Identifier) ist ein Experiment, das erst längerfristig in größerem Umfang Früchte tragen wird. Je kooperativer katalogisiert wird, desto kooperativer kann auch von Anreicherungen profitiert werden. D.h. wenn in der Formalerschließung konsequent verbundübergreifend Fremddaten genutzt werden, dann ist es möglich, Korrekturen und Ergänzungen aus den Bereichen Formalerschließung und auch Sacherschließung aus anderen Verbänden und von der DNB mit einfachen maschinellen Update-Verfahren ohne manuelles Eingreifen zu übernehmen.

Als Aufgabe bleibt, sich der bestehenden Differenzen bei Regeln, Formaten und Qualitätsstandards anzunehmen und hier zu pragmatischen Lösungen zu kommen. Es ist allgemein bekannt, dass eben diese Fragen zurzeit an unterschiedlichen Stellen diskutiert werden: im Zusammenhang mit der Kooperationsinitiative hzb – BVB – OBV, dem Formatumstieg MAB – MARC 21 und nicht zuletzt der RDA-Entwicklung.⁷ Da im ersten Halbjahr 2007 mit konkreten Ergebnissen gerechnet werden kann,⁸ wird die Projektgruppe „Kooperative Neukatalogisierung“ ihre Arbeit an diesem Bereich Mitte 2007 aufnehmen und überlegen, inwieweit sich aus der Praxis vermehrter gegenseitiger Datenübernahmen

⁷ Durch personelle Überschneidungen zwischen den hier aktiven Arbeits- und Expertengruppen und der Projektgruppe „Kooperative Neukatalogisierung“ ist der Informationsfluss gewährleistet und es kommt nicht zu Doppelarbeit.

⁸ Dies gilt z.B. für den Bereich des Austauschformates. Hier werden von der MARBI-Frühjahrssitzung wichtige Entscheidungen für die zukünftige Entwicklung des MARC-Formates erwartet.

pragmatische Lösungen finden lassen. Der Formatumstieg auf MARC 21 bedeutet für alle Projektpartner in gewissem Umfang einen Neubeginn. Dies bietet auch die Chance, durch Festlegungen zum Format im Zuge des Umstiegs die Entwicklung von verbundspezifischen Format-Dialekten zu verhindern.