

Electronic Resource Management für alle? Automatisierte Contentanalyse mit Excel und VBA

André Schüller-Zwierlein

Electronic Resource Management für alle?

Seit die Electronic Resource Management Initiative der Digital Library Foundation (DLF ERMI) im Jahre 2004 ihren Abschlussbericht veröffentlicht hat,¹ ist das Thema Electronic Resource Management (ERM) in aller Munde. Die Vision der DLF-Initiative, den gesamten Lebenszyklus eines elektronischen Produktes in einer Bibliothek in einem System zu verwalten – von der Contentanalyse vor dem Kauf bis hin zur Verwaltung von Lizenzvertragsdaten und der automatischen Lieferung von COUNTER-gerechten Nutzungsstatistiken (die nun über das Protokoll SUSHI möglich geworden ist) –, ist in verschiedene kommerziell vertriebene Systeme umgesetzt worden, die äußerst leistungsfähig sind. Bei den großen in Europa vertriebenen Systemen wie Verde (ExLibris) und ERMS (ProQuest) scheint sich jedoch herauszukristallisieren, dass sie für eine einzelne Bibliothek, obwohl ursprünglich für diese intendiert, kaum finanzierbar sind. Dementsprechend wird der Einsatz dieser umfassenden Lösungen momentan eher durch Konsortialgeschäftsstellen und Verbände erwogen: die HEBIS-Zentrale setzt z.B. bereits ein ERM-System ein,² beim GBV hat sich eine AG ERM³ gebildet, die zunächst die Anforderungen an ein solches System erörtert.

Es gibt jedoch auch Produkte, die als Alternativen oder Ergänzungen zu diesen großen Systemen aufzufassen sind und die Teillösungen für den ERM-Bereich erbringen: Ein Beispiel ist ScholarlyStats (MPS Technologies; Vertrieb: Swets),⁴ ein Portal, das für die einzelne Bibliothek COUNTER-gerechte Nutzungsstatistiken für E-Zeitschriften und Datenbanken großer Plattformen/Hosts bietet – im Endeffekt ein Outsourcing der Sammelarbeit, die durch die verschieden formatierten und an verschiedenen Orten angebotenen Statistikdaten der einzelnen Hersteller entsteht. Eine andere kommerzielle Teillösung ist das Ulrich's Serials Analysis System (USAS) (CSA),⁵ das auf der Basis der Ulrichsweb-Daten einen Vergleich von Content-Listen im Hinblick auf die Produktbewertung, z.B. von Aggregatordatenbanken, ermöglicht. Auch bei diesen Produkten stellt sich allerdings die Frage, ob

1 <http://www.diglib.org/pubs/dlf102/ERMFINAL.pdf>

2 http://www.hebis.de/bib/veranstaltungen/HeBIS-Nutzer_060913.pdf

3 http://www.gbv.de/wikis/cls/AG_ERM

4 <https://www.scholarlystats.com/sstats/default.htm>

5 <http://www.ulrichsweb.com/ulrichsweb/analysis/>

ein jährlich zu entrichtender vierstelliger Betrag für eine durchschnittliche Universitätsbibliothek eine realistische Option darstellt.

Daher wird immer mehr über Alternativlösungen nachgedacht, um alltägliche Prozesse in diesem Bereich auch für die einzelne Bibliothek zu erleichtern. Auf einem in Zusammenarbeit mit der Expertengruppe Erwerbung und Bestandsentwicklung des DBV⁶ veranstalteten Workshop der Zentralbibliothek des Forschungszentrums Jülich z.B. wurden im Herbst 2006 selbst erstellte ERM-Lösungen aus Jülich und Ilmenau vorgestellt. Aber vor allem für Teilfunktionen eines ERM-Systems lassen sich auf der Basis gängiger Software vergleichsweise einfach selbst Programme erstellen, die regelmäßig zu wiederholende Arbeitsprozesse automatisieren und so Arbeitszeit sparen. Dies soll im folgenden am Beispiel einer in VBA programmierten Excel-Lösung für die automatisierte Contentanalyse, z.B. bei Aggregatordatenbanken, demonstriert werden.

Automatisierte Contentanalyse

Für Fachreferenten wie Erwerbungsleiter stellen sich immer wieder dieselben Fragen: Wie gut ergänzt eine neu angebotene Aggregatordatenbank das Volltextangebot der Bibliothek, wie weit überlappen sich zwei Angebote, rechtfertigt die Anzahl singulärer Titel einer Datenbank einen Kauf? Oft werden in deutschen Bibliotheken solch zentrale Fragen aus Zeit- oder anderen Gründen ‚aus dem Bauch heraus‘ bzw. nach einem cursorischen Vergleich entschieden; eine methodische Auswertung der Inhalte findet allzu selten statt. Eine Lösung für diese Fragen bietet z.B. das oben genannte USAS:⁷ Zu dessen Vorzügen gehören die aktuellen und einheitlich formatierten Daten aus Ulrichsweb sowie die Knowledge Base gängiger Aggregatordatenbanken. Es lässt sich jedoch auch eine kostengünstigere Lösung finden: Setzt man voraus, dass die Hersteller von Aggregatordatenbanken jeweils aktuelle Contentlisten als Excel-Datei (oder mindestens in importierbarem Format) auf ihren Homepages anbieten, und ist man bereit, die wenigen Handgriffe vorzunehmen, die evtl. zur Anpassung der Formatierung erforderlich sind, so kann der systematische Contentvergleich nach ISSN (auf Wunsch auch nach Titel) mit Angabe der Mengen singulärer und dubletter Titel sowie Prozentzahlen auch durch eine in VBA programmierte Excel-Lösung möglich gemacht werden.

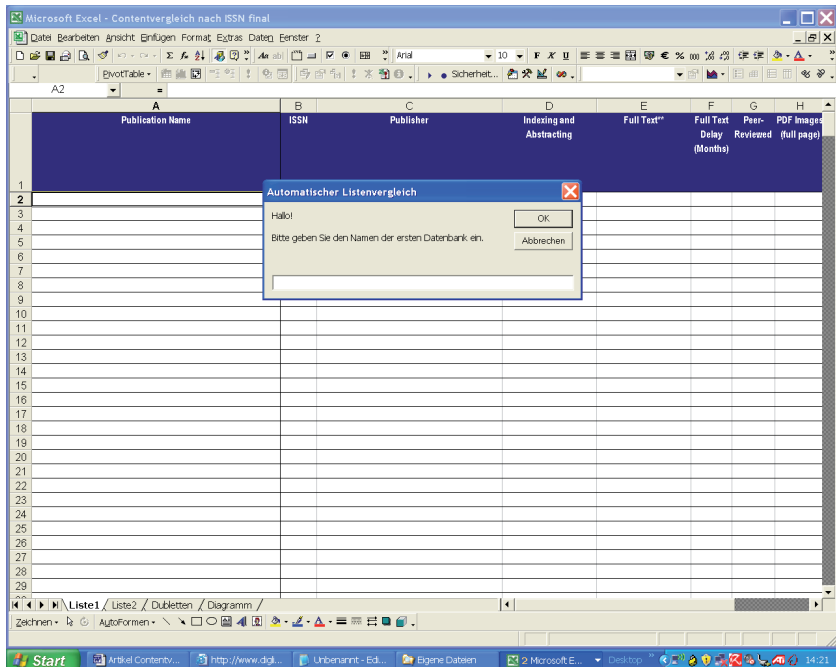
Das hier vorgestellte Tool steht unter <http://www.ub.uni-muenchen.de/pdfs/Contentvergleich.xls> zum kostenlosen Download bereit. (Nach dem Anwählen des Links sollten die Makros deaktiviert werden, dann kann die Datei einfach ab-

6 http://www.fz-juelich.de/zb/Programm_Okt_2006/

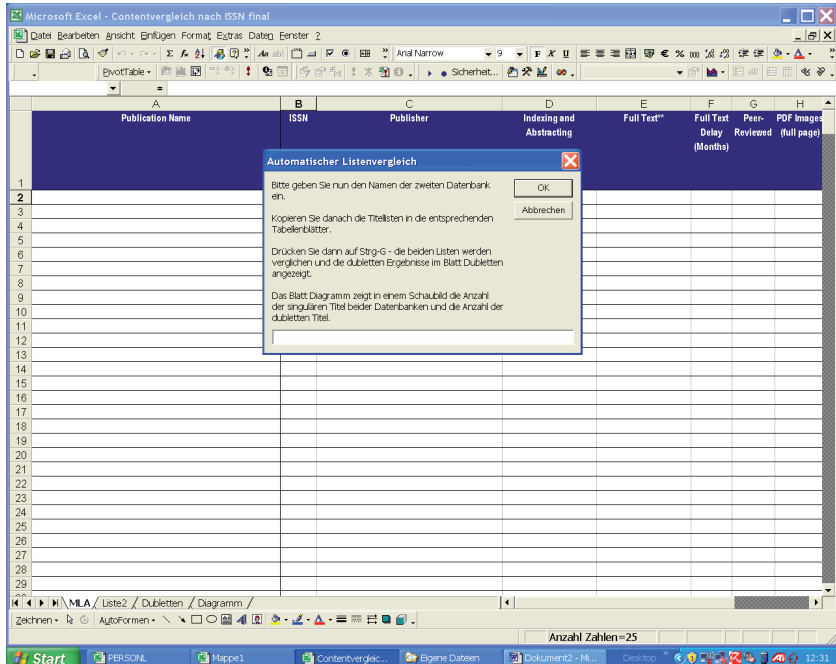
7 Zu den Einsatzmöglichkeiten s. z.B. Alice Keller; Dag Martinsen: "Serials rationalization at Oxford University library services using Ulrich's serials analysis system". *Serials* 18 (2005) 1, 38–44.

gespeichert werden.) Das Tool steht frei zur lokalen Anpassung und Veränderung zur Verfügung.

Damit das Tool ordnungsgemäß funktioniert, sollte die Makrosicherheit in Excel (Extras/Makro/Sicherheit) auf ‚Mittel‘ stehen. Beim Öffnen der Excel-Datei wird zunächst gefragt, ob die Makros aktiviert werden sollen – dies ist zu bestätigen. Dann wird der Name der ersten Datenbank abgefragt. Beim Vergleich anderer Listen kann ein beliebiger Name eingegeben werden.



Danach wird der Name der zweiten Datenbank abgefragt. Die beiden Tabellenblätter „Liste1“ und „Liste2“ werden automatisch entsprechend umbenannt.



Nun gilt es die Titellisten in die zwei den Datenbanken entsprechenden Tabellenblätter zu kopieren. Die aufgelisteten Kriterien/Spalten entsprechen gängigen Titellisten großer Hersteller.⁸

⁸ Ist ein anderes Format gewünscht, muss ggf. der VBA-Code angepasst werden.

Microsoft Excel - Contentvergleich nach ISSN final

	A	B	C	D	E	F	G	H
	Publication Name	ISSN	Publisher	Indexing and Abstracting	Full Text**	Full Text Delay (Months)	Reviewed	Peer-Reviewed (full page)
1								
2	AANA Journal	10624730	American Association of Nurse Anesthetists	01/01/2001 to present	01/01/2001 to present		✓	✓
3	ABA Journal	10610057	American Bar Association	01/01/1975 to present	01/01/1975 to present		✓	✓
4	Abdominal Imaging	0965880X	Springer-Verlag	06/01/2003 to present			✓	
5	ABNF Journal	0956182X	Tucker Publications, Inc.	11/01/2002 to present	11/01/2002 to present		✓	✓
6	About Campus	01640178	John Wiley & Sons, Inc.	01/01/1997 to present	01/01/1997 to present	12	✓	✓
7	Academic	02777126	American Association of University Professors	01/01/1992 to present	01/01/2002 to present		✓	
8	Academic Leader	09087146	Magna Publishers Inc.	01/01/2004 to present	01/01/2004 to present		✓	
9	Academic Questions	02633709	Transaction Publishers	12/01/1987 to present	12/01/1987 to present	6	✓	✓
10	ACC Current Journal Review	00030978	Elsevier Science	01/01/2002 to present			✓	✓
11	Accent on Living	00030907	Creever Publishing, Inc.	01/01/1990 to 01/01/2001	01/01/1990 to 01/01/2001		✓	✓
12	Access	01439017	American Dental Hygienists Association	01/01/2001 to present	01/01/2001 to present		✓	✓
13	Accident Analysis & Prevention	0952769X	Pergamon Press - An Imprint of Elsevier Science	01/01/2002 to present			✓	
14	Accounts of Chemical Research	00035769	American Chemical Society	04/01/2003 to present			✓	
15	ACES Bulletin	00041610	Association for Comparative Economic Studies	03/01/1972 to 12/01/1984			✓	
16	ACM Computing Surveys**	07425427	Association for Computing Machinery	01/01/1971 to present			✓	
17	ACM Transactions on Computer Systems	01624962	Association for Computing Machinery	06/01/1996 to present			✓	
18	ACM Transactions on Database Systems**	10393114	Association for Computing Machinery	01/01/1996 to present			✓	
19	ACM Transactions on Information Systems	00984246	Association for Computing Machinery	06/01/1996 to present			✓	
20	ACM Transactions on Mathematical Software	00070904	Association for Computing Machinery	03/01/1996 to present			✓	
21	ACM Transactions on Programming Languages & Systemsx	00072869	Association for Computing Machinery	03/01/1996 to present			✓	
22	Acoustical Physics	01612492	MAK Nauka Interperiodica	01/01/2000 to present	01/01/2000 to present	12	✓	✓
23	ACP Journal Club	0001190X	American College of Physicians	01/01/2002 to present	01/01/2002 to present		✓	
24	Acta Agriculturae Scandinavica: Section A, Animal Science	00014360	Taylor & Francis Ltd	02/01/1998 to present	02/01/1998 to present	12	✓	✓
25	Acta Agriculturae Scandinavica: Section B, Soil & Plant	00016456	Taylor & Francis Ltd	03/01/1998 to present	02/01/1998 to present	12	✓	✓
26	Acta Agriculturae Scandinavica: Section C - Food Economics	00039396	Taylor & Francis Ltd	04/01/2004 to present	04/01/2004 to present	12	✓	✓
27	Acta Anaesthesiologica Scandinavica	09020205	Elsevier Publishing Limited	01/01/1999 to present	01/01/1999 to present	12	✓	✓
28	Acta Archaeologica	09342043	Bisshel Publishing Limited	11/01/2001 to present	11/01/2001 to present	12	✓	✓

Sind beide Listen in das jeweilige Tabellenblatt eingefügt, kann die Vergleichsfunktion mit Strg-G gestartet werden.

The screenshot shows an Excel spreadsheet with columns: Publication Name, ISSN, Publisher, Indexing and Abstracting, Full Text*, Full Text Delay (Months), Peer-Reviewed, and PDF Images (full page). A dialog box titled 'Ergebnis' is open, displaying the following text:

30 Titel sind in beiden Datenbanken vorhanden.
 37,5 Prozent der Titel aus der Datenbank MLA sind auch in der Datenbank ASP vorhanden.
 54,5 Prozent der Titel aus der Datenbank ASP sind auch in der Datenbank MLA vorhanden.
 Sie können sich nun die Details im Blatt Dubletten ansehen.
 Das Blatt Diagramm zeigt in einem Schaubild die Anzahl der singulären Titel beider Datenbanken und die Anzahl der dubletten Titel.

Buttons: OK

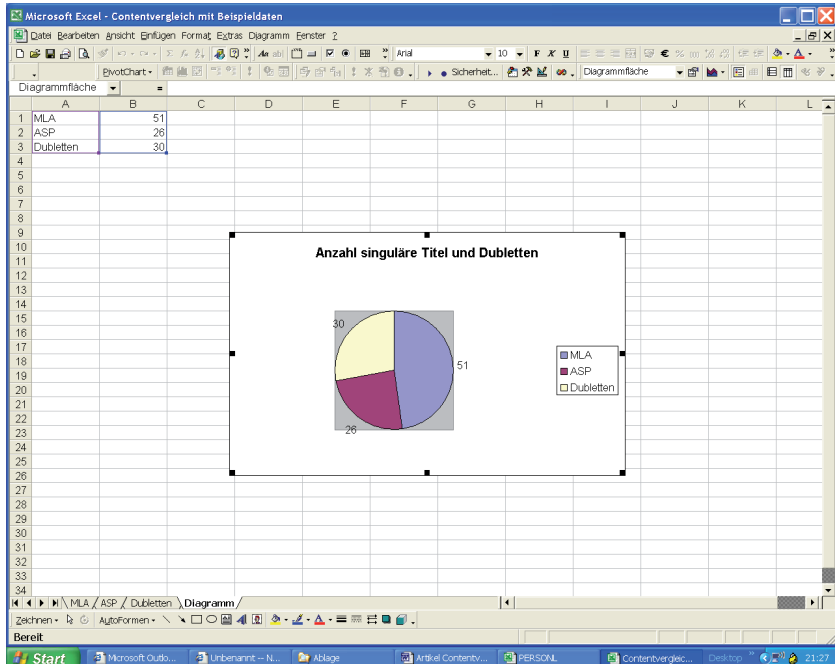
Als Ergebnis wird die Anzahl der dubletten Titel (nach ISSN) ausgegeben, daneben der prozentuale Anteil von Dubletten an der jeweiligen Datenbank. Die dubletten Titel mitsamt allen Daten werden im Tabellenblatt „Dubletten“ ausgegeben – so kann verglichen werden, mit welchem Berichtszeitraum, welcher Volltextart, wievielen Volltexten etc. der dublette Titel in der jeweiligen Datenbank vertreten ist.

	A	B	C	D	E	F	G	H
	Publication Name	ISSN	Publisher	Indexing and Abstracting	Full Text*	Full Text Delay (Months)	Peer-Reviewed	PDF Images (full page)
1								
2	Acta Odontologica Scandinavica	00151386	Taylor & Francis Ltd	04/02/1998 to present	04/02/1998 to present	12	✓	✓
3	Test 1	00151386	American Association of Nurse Anesthetists	01/01/2000 to present	01/01/2000 to present		✓	✓
4	Acta Odologica	0015587X	Elsevier Science	01/01/2002 to present			✓	
5	Test 2	0015587X	American Bar Association	01/01/1975 to present	01/01/1975 to present		✓	
6	Acta Oncologica	00151818	Taylor & Francis Ltd	05/01/1998 to present	05/01/1998 to present	12	✓	✓
7	Test 3	00151818	Springer-Verlag	05/01/2002 to present			✓	

Um sehen zu können, welche Titel in der jeweiligen Datenbank singular sind und damit ihren Mehrwert zu ermitteln, werden die dubletten Titel in den ursprünglichen Listen rot markiert.

	A	B	C	D	E	F	G	H
	Publication Name	ISSN	Publisher	Indexing and Abstracting	Full Text**	Full Text Delay (Months)	Peer-Reviewed	PDF Images (full page)
1								
54	Ad Astra	0163075X	National Space Society	01/01/1989 to present				
55	About Campus	01640178	John Wiley & Sons, Inc.	01/01/1997 to present	01/01/1997 to present	12	✓	✓
56	Addiction Research	02691205	Taylor & Francis Ltd	04/01/1997 to 12/01/2000	04/01/1997 to 12/01/2000		✓	
57	Acta Oto-Laryngologica	02730340	Taylor & Francis Ltd	05/27/1999 to present	05/27/1999 to present	12	✓	✓
58	Acta Orthopaedica Scandinavica	02763382	Taylor & Francis Ltd	03/01/1998 to present	03/01/1998 to present	12	✓	✓
59	Academex	02777126	American Association of University Professors	01/01/1992 to present	01/01/1992 to present			
60	Addition Research & Theory	02739979	Elsevier Publishing	01/05/2001 to present	01/01/2001 to present	6	✓	
61	Acta Haematologica	03071762	Karger AG	01/01/1998 to present			✓	
62	Acta Neurochirurgica	03160300	Kluwer Academic Publishing / Academic	04/01/2003 to present			✓	
63	Academic Questions	03603700	Transaction Publishers	12/01/1997 to present	12/01/1997 to present	6	✓	✓
64	Acta Zoologica Lituonica	03636941	Institute of Ecology Vilnius University	03/01/2002 to present	03/01/2002 to present		✓	
65	ADHD Report	07211403	Gulford Publications Inc.	01/01/2001 to present				
66	Acta Radiologica	0738340X	Taylor & Francis Ltd	01/01/1999 to present	01/01/1999 to present	12	✓	✓
67	ACM Computing Surveys**	07426457	Association for Computing Machinery	01/01/1971 to present			✓	
68	Acta Histochemica	07426473	Urban & Fischer Verlag	04/01/2002 to present			✓	
69	Addition	07482926	Blackwell Publishing Limited	03/01/1999 to present	03/01/1999 to present	12	✓	✓
70	Acta Archaeologica	08182043	Blackwell Publishing Limited	11/01/2001 to present	11/01/2001 to present	12	✓	✓
71	Acta Neuropsychiatrica	08482449	Blackwell Publishing Limited	01/01/2003 to present	01/01/2003 to present	12	✓	✓
72	Acta Psychologica	08752522	Elsevier Science	01/01/2002 to present			✓	
73	Accident Analysis & Prevention	0896760X	Pergamon Press - An imprint of Elsevier Science	01/01/2002 to present			✓	
74	Academic Leader	08967148	Magna Publications Inc.	01/01/2004 to present	01/01/2004 to present			
75	Acta Paediatrica	0905660X	Taylor & Francis Ltd	01/05/1998 to present	01/05/1998 to present	12	✓	✓
76	Acta Mathematica Sinica	10596370	Springer-Verlag	01/01/2000 to present	01/01/2000 to present	12	✓	
77	ABA Journal	10610057	American Bar Association	01/01/1975 to present	01/01/1975 to present		✓	
78	Acta Ophthalmologica Scandinavica	10775397	Blackwell Publishing Limited	01/01/2000 to present	01/01/2000 to present	12	✓	
79	ACM Transactions on Database Systems**	10831741	Association for Computing Machinery	01/01/1986 to present			✓	
80	Adaptive Behavior	10836885	Gaige Publications	01/01/1998 to present			✓	
81	Acta Physiologica Scandinavica	15412293	Blackwell Publishing Limited	01/01/1999 to present	01/01/1999 to present	12	✓	✓

Zusätzlich wird die Anzahl der jeweils singulären und der dubletten Titel im Blatt „DiagrammV in einem Schaubild ausgegeben.



Damit sind grundlegende Daten für die Analyse und Bewertung der Produkte zusammengestellt. Die Datei sollte nun unter einem neuen Namen gespeichert werden.

Fazit

Obwohl in deutschen Bibliotheken in sehr vielen Fällen die gleichen Programme zur Erledigung täglicher Aufgaben verwendet werden, ist der Austausch kleiner EDV-Lösungen und bibliothekarischer Tools unterentwickelt. Hier sollte evtl. eine entsprechende Plattform geschaffen werden. Gerade angesichts meist sehr unterschiedlicher EDV-Fertigkeiten der Mitarbeiter/innen in den Bibliotheken erscheint eine durchdachte Automatisierung oft wiederholter Arbeitsprozesse wünschenswert. Die umfassenden Möglichkeiten, die z.B. das Office-System hierzu bietet, über VBA und in Zukunft verstärkt über VB.NET und C#, werden immer noch unzureichend genutzt. Besonders bei Statistik-/Reporting-Funktionen und bei der Datenanalyse gilt es gründlich zu prüfen, ob die Leistungen kommerziell angebotener Produkte nicht auch mit einfachen Mitteln und durch gründliche Ausnutzung der Funktionsvielfalt gängiger Software erbracht werden können.

Aus der 51. Sitzung der Arbeitsgemeinschaft der Verbundsysteme am 29. und 30. November 2006 in Frankfurt am Main

Susanne Oehlschläger

Am 29. und 30. November 2006 hat die Arbeitsgemeinschaft der Verbundsysteme auf Einladung der HeBIS-Verbundzentrale ihre 51. Sitzung in der Universitätsbibliothek Johann Christian Senckenberg in Frankfurt am Main durchgeführt.

Zusammenarbeit der Verbundsysteme, Verbesserung von Datenausch und Fremddatennutzung

Zur Umsetzung der Beschlüsse des Standardisierungsausschusses zur Internationalisierung des Regelwerkes und der Formate insbesondere durch die Verbesserung der Zusammenarbeit der Verbundsysteme untereinander sowie des Datenauschs und der Fremddatennutzung hat die Arbeitsgemeinschaft der Verbundsysteme verschiedene Projekte angestoßen. Dazu gehören u.a. die Projekte „Umstieg auf MARC 21“, „Kooperative Neukatalogisierung“ und „Catalogue Enrichment“.

Umstieg auf MARC 21

Zur Vorbereitung des Umstiegs auf MARC 21 haben die Mitglieder der Expertengruppe Datenformate im September 2006 gemeinsam mit den Spezialisten der Deutschen Nationalbibliothek und der Leiterin der Abteilung Network Development and MARC Standards an der Library of Congress, Sally McCallum, im Rahmen eines Workshops an der Deutschen Nationalbibliothek die Konkordanz von MAB2 auf MARC 21 sowie offene Punkte diskutiert und Lösungswege erarbeitet. Im Anschluss daran wurden eine Gesamtdarstellung der Lösungskonzeption (Discussion Paper) und detaillierte Vorschläge („proposals“) für die Einrichtung neuer Felder erarbeitet. Bei der Sitzung des Machine Readable Bibliographic Information Committee (MARBI) Ende Januar 2007 in Seattle, Wash. wird Reinhold Heuvelmann (DNB) das zuvor erarbeitete Discussion Paper vorstellen. Adrian Nolte (Stadtbibliothek Essen für die Öffentlichen Bibliothekssysteme) nimmt als Vertreter der Expertengruppe Datenformate ebenfalls an der Sitzung teil. Als nächste Aktivitäten sind ein Workshop mit internationaler Beteiligung im Mai und ein Workshop mit den Systemanbietern im Sommer 2007 geplant.

Entgegen der ursprünglichen Absicht, eine eigene Arbeitsgruppe für Planung und Koordinierung des konkreten Umstiegs zu gründen, soll dieser von der Arbeitsgemeinschaft der Verbundsysteme und ggf. den vorhandenen Gruppen