Ein Kessel Buntes 2002
Sacherschließung im SWB

Stefan Wolf, BSZ Konstanz

Die Behauptung steht im Raum, wir könnten alles in Baden-Württemberg – natürlich außer Hochdeutsch. Dann gehört bibliothekarische Sacherschließung sicher zu unseren Kompetenzfeldern. Um die Richtigkeit der Annahme zu belegen, sei im Folgenden ausgeführt, was im Bibliotheksservice-Zentrum Baden-Württemberg (BSZ) und auf der Verbunddatenbank des Südwestdeutschen Bibliotheksverbundes (SWB) im Lauf des Jahres 2002 geschehen ist oder gerade realisiert wird.

Vorgestellt werden einzelne Projekte, die zunächst scheinbar unverbunden sind. Wenn sich am Ende herausstellt, dass Sacherschließung im Verbund bedeutet, unter Optimierung von Werkzeugen und Geschäftsgängen die Erschließungsdichte und -tiefe zu verbessern, gleichzeitig aber verschiedene Systeme zu nutzen, so wird erkennbar, dass bibliothekarische Sacherschließung nicht ein einziges monolithisches Arbeitsfeld mit der Geschlossenheit der Formalerschließung darstellt. Vielmehr werden, um den vorhandenen Daten und den Bedürfnissen der Nutzerschaft in Recherche und Auskunft gerecht zu werden, verschiedene Datenquellen angeboten, wo möglich verbunden und genutzt werden müssen. Im Verbund wird – um die Ausgangsthese aufzugreifen – nicht nur hochdeutsch gesprochen, sondern eben z.B. auch "altdeutsch" mit den Museen und Archiven, "bayerisch" mit der Regensburger Verbundklassifikation oder amerikanisch mit der Dewey Decimal Classification.

Eine gemeinsame Sprache mit Archiven und Museen...
Intensiv und zum Teil kontrovers wurde im deutschen Museumswesen in den vergangenen Jahren diskutiert, ob bibliothekarische Werkzeuge wie die SWD in der Museumsdokumentation sachgerecht verwendet werden können. Die Debatte ist zu einem Teil akademisch, da einige Einrichtungen wie z.B. das Landesmuseum für Technik und Arbeit in Mannheim dies seit Jahren mit gutem Erfolg und großer Zufriedenheit praktizieren. Sowohl durch das MusIS- wie auch das BAM-Projekt (Einführung einer integrierten Museumsverwaltungssoftware bzw. Aufbau eines gemeinsamen Internet-Portals für Bibliotheken, Archive und Museen) kam ein zusätzlicher Impuls: die Nutzung der SWD alleine genügt auf Dauer nicht; sowohl eine technische wie auch eine redaktionelle Infrastruktur muss aufgebaut werden, um neue Deskriptoren und Korrekturvorschläge, die in diesen Institutionen notwendig werden, in die SWD einführen zu können.

Es wurde deshalb ein Projekt zunächst zwischen Der Deutschen Bibliothek (DDB), BSZ, Landesmuseum für Technik und Arbeit Mannheim, Badischem Landesmuseum Karlsruhe und Landesarchivdirektion Stuttgart verabredet: Über vom BSZ zu entwickelnde Erfassungsmasken werden neue Deskriptoren erfasst und neben der Anwendung "SWD im WWW" des BSZ zwischengespeichert. Mit Hilfe von zu entwickelnden Skripten werden diese Schlagworte in das PICA/Iltis-Erfassungsformat konvertiert und direkt einzeln im Katalogisierungssystem Der Deutschen Bibliothek abgelegt. Termingerecht waren Erfassungsmasken, Konversionsskripte und auch die notwendigen Redaktions- und Kommunikationsstrukturen neben den bibliothekarischen Verfahren abgesprochen bzw. fertig gestellt. Der konzipierte Test von 6 Monaten (Januar – Juni 2002) war so erfolgreich, dass sowohl der Einbezug weiterer Partner wie auch die Fortsetzung problemlos gelangen. Dank gebührt Der Deutschen Bibliothek sowohl für das freundliche Eingehen auf ein solches Begehren wie auch die kollegiale Übernahme redaktioneller Aufgaben besonders im Testzeitraum. Nicht zuletzt die SWD profitiert aber davon, dass neue Deskriptoren sachgerecht eingeführt wurden, zu denen bislang eigenständige Literatur nicht erschienen ist

(z.B. Salatbesteck oder bestimmte Kleidungsteile), die aber selbstverständliche und wertvolle Sammlungsgegenstände in Museen aller Art und Größe sind.

...beschleunigt den bibliothekarischen Geschäftsgang
Das etablierte Redaktionsverfahren neuer SWD-Termini profitiert von diesen Werkzeugen: Verabredet wurde mit der Zentralredaktion Schlagworte des SWB an der UB Kaiserslautern auf der Grundlage der für die Archive und Museen entwickelten Skripte, dass die bislang manuelle Übergabe neuer Schlagwortsätze (ca. 5.000 jährlich) in Zukunft teilautomatisiert erfolgen wird.

In KATWIN wird ein MAB2-Download einzeln auszuwählender Schlagwortsätze in eine gemeinsame Datei realisiert. Dieser Wunsch der Kolleginnen und Kollegen in Kaiserslautern wird weitergeführt, indem diese Datei dann für den Import in PICA/Iltis Der Deutschen Bibliothek verwendet wird. Unter Verwendung der für die Archiv- und Museumskooperation entwickelten Konversionsskripte wird die gleiche Übergabe-Funktionalität in bzw. an PICA/Iltis geschaffen: Der einzelne Satz kann direkt aus der MAB2-Datei in diesem Erfassungsformat aufgerufen, bearbeitet und abgespeichert werden. Parallel wird in eine zweite Datei nach der Speicherung in PICA/Iltis die ID in SWD und SWB festgehalten, die wiederum in die betroffenen Sätze eingetragen werden kann.

Gemeinsam erwarten wir eine komfortablere und beschleunigte Bearbeitung und Übergabe neuer Schlagwortsätze an die SWD bei Der Deutschen Bibliothek, die ohne umständliche Copy&Paste-Prozeduren auskommt und eine zeitlich weniger verzögerte Übergabe an die DDB ermöglicht, womit ein altes Anliegen vieler lokaler Redaktionen eingelöst würde.

(Jetzt im Jan. 2003 während der Redaktion der schriftlichen Fassung des Vortragstextes befindet sich dieses Werkzeug bereits im ersten Einsatz.)

Zur aktuellen Situation der Sacherschließung im SWB
Die verbale kooperative Sacherschließung nach RSWK unter Nutzung der SWD und der Sacherschließungsdaten Der Deutschen Bibliothek läuft im Routinebetrieb. Erreicht wurde mittlerweile, dass an ca. 1,3 Mio. Titelaufnahmen etwa 1,6 Mio. Ketten vorhanden sind – angesichts der zahlreichen Konversionsprojekte, die keine regional nutzbare Sachinformation lieferten, eine vertretbare, aber nicht wirklich befriedigende Quote von ca. 15 Prozent der Aufnahmen im SWB. Dass daneben nach eigenen Systemen vielfältige Wege in den Lokaldaten beschritten wurden und werden, ist bekannt. Nach und nach etabliert sich die RVK als gemeinsam genutzte klassifikatorische Komponente nach dem Beispiel des Bayerischen Bibliotheksverbunds (BVB); eine beachtliche Anzahl von Teilnehmereinrichtungen systematisiert aber immer noch über die Lokaldaten.

Wünschenswert ist also insgesamt eine breitere Teilnahme an der Kooperation – sowohl die Zahl der Häuser, als auch das Spektrum der Fachgebiete betreffend. Um dies zu unterstützen, hat das BSZ im Jahr 2002 mehrere Schulungen organisiert, die auf Wunsch wiederholt werden können.

Desiderat ist genauso ein insgesamt höherer Anteil der direkt im Titelstamm verbal und klassifikatorisch erschlossenen Aufnahmen. Klar ist, dass eine breit angelegte aktive retrospektive Sacherschließung aus Aufwandsgründen von vornherein ausscheidet. Überlegt wurde, woher große, der Struktur und dem derzeitigen System entsprechende Datenmengen bezogen werden könnten. Naheliegend war, auf den Bayerischen Bibliotheksverbund zuzugehen, um Möglichkeiten des Datentauschs auszuloten angesichts der bekanntermaßen hohen sachlichen Erschließungsquote dort.

Übernahme der Sacherschließungsinformation aus dem BVB
Im Rahmen der Kooperation zwischen den Bibliotheken Bayerns, Baden-Württembergs und Sachsens wurde das gegenseitige Interesse erklärt, ein solches Projekt durchzuführen. Die Durchführung wurde in die Wege geleitet, indem dem BVB ein Gesamtabzug der SWB-Titeldaten übergeben wurde, der aus zwei Teilen bestand: eine Tranche bestand aus Titelaufnahmen, die bereits verbal sachlich erschlossen waren. Sie sollten nicht nochmals mit RSWK-Ketten angereichert werden; lediglich die Klassifikationsinformation war gesucht. Die zweite, beträchtlich größere Tranche bestand aus den Titelaufnahmen, die noch über keine verbale Indexierung verfügte. Dankenswerterweise übernahmen die Kolleginnen und Kollegen vom BVB den automatisierten Titelabgleich, also das Matching der Titelaufnahmen. Als Ergebnis spielten sie die vorhandene sachliche Erschließungsinformation den Titelaufnahmen des SWB zu und lieferten die Daten zurück an die Verbundzentrale des SWB.
Automatisch identifiziert wurden so knapp 2 Mio. Titelaufnahmen, die nun mit neu zu entwickelnden Programmen zu bearbeiten waren.

Bemerkt sei, dass als Gegenleistung die gesamte Autorenstammdatei des SWB dem BVB für den Aufbau einer Personenstammdatei zur Verfügung gestellt wurde.

... von Notationen der Regensburger Verbundklassifikation
Die Analyse der Daten und Redaktionskonzepte besonders in Zusammenarbeit mit der SLUB Dresden zeigte, dass die Einspielung der verwendeten RVK-Notationen einfach zu bewerkstelligen war: unter Verzicht auf den Abgleich mit der Normdatei wurden in den Kategorien 720ff im Mai und Juni 2002 zu 441.000 Titeln (eben denjenigen, die bereits SWD-Ketten besassen) 721.000 Systemstellen eingetragen. Dabei wurde, falls im Rahmen der RVK-Kooperation im SWB noch kein eigener Datensatz in der regionalen Notationsstammdatei für die einzelne Systemstelle vorhanden war, ein neuer Satz angelegt und über ein Abrufzeichen gekennzeichnet. Tests zeigten, dass durchaus verschiedene Redaktionsstände der RVK ihren Niederschlag gefunden hatten. Die Zahl der Ausreißer war nicht so groß, dass das Verfahren als untauglich hätte charakterisiert werden müssen; vereinbart wurde, die Sätze in jedem Fall im Nachhinein einer Sichtung zu unterwerfen, um eine redaktionell hochwertige RVK-Systematik im SWB zu bewahren, die dem Anspruch einer unterlegten Normdatei genügt.

... von Schlagwortketten nach RSWK und SWD
Ungleich schwieriger gestaltete sich die Ausgangssituation bei der anstehenden Übernahme der Schlagwortinformation. Zwar ist der BVB seit Anbeginn Kooperationspartner beim Entstehen der SWD, führt diese aber bislang nicht als verknüpfte Normdatei. Geliefert werden konnte deshalb vom BVB nur der reine Ansetzungstext; Identnummern, die Schlagwortart kennzeichnende Steuerbuchstaben fehlten ebenso wie ein unterscheidendes Merkmal als Trennzeichen in Ansetzungs- und Verknüpfungsketten.

Geliefert wurde also in MAB2 eigentlich eine bloße Textauflistung – die z.B. so aussehen konnte:

902 USA
902 Verkehrspolitik
902 Eisenbahn
902 Güterverkehr

Ob dieser Text vier einzelne SWD-Termini darstellt oder eine Ansetzungskette enthält (z.B. USA / Verkehrspolitik), war den Daten nicht zu entnehmen. Ein Programm war zu entwickeln, das den gelieferten Text unter fachgerechtem, aber vielfachem Abgleich an der Schlagwortstammdatei des SWB analysiert und zuordnet.

Der Grundalgorithmus wurde so gestaltet, dass zunächst versucht wurde, den ganzen Text als eine Schlagwortansetzung abzugleichen. Konnte diese nicht gefunden werden, wurde das letzte Glied abgetrennt, der Rest und getrennt davon das bereits separierte letzte Glied untersucht. Je nach Länge der Kette wurde so eine wachsende Zahl von Anfragen auf der Datenbank abgewickelt. Als Ergebnis konnte eine erstaunlich gute Identifizierung von Schlagworttext und Schlagwortstammdatei erreicht werden. Die im Vorfeld der Projektdurchführung diskutierte Frage stellte sich nicht mehr wirklich, ob eine Verknüpfung mit der Schlagwortstammdatei überhaupt erreicht werden kann, ohne deren Qualität nachhaltig zu beeinträchtigen.

Weitere Unterprogramme waren nötig: so mussten Benutze-Kombination-Hinweise (auch "Eigentlich Verknüpfungsketten") identifiziert und zur Verknüpfung mit den eigentlich bezeichneten Einzelschlagworten aufbereitet werden. Schlagwörter mit Zeitangaben mussten ihrer Gattung (Form-, Sach-, Zeitschlagwörter) zugeordnet werden. Dies gelang gut dank der RSWK-inhärenten Systematik der Ansetzung dieser Sachverhalte. Die Unterscheidung von gleich angesetzten, nur durch den Steuerbuchstaben unterschiedenen Form- und Sachschlagworten wurde im Regelfall
a priori zugunsten des Formschlagwortes entschieden. In einigen Fällen war es aber sinnvoll, die Voreinstellung zugunsten des Sachschlagworts vorzunehmen. Die Kennzeichnung neu entstandener Schlagwortverwendungen war festzulegen, ebenso wie eine sachgerechte Protokollierung und Statistik des komplexen Gesamtablaufs.

Ein erster Test mit sehr detaillierter Überprüfung des Ergebnisses in der Verbundzentrale Anfang August 2002 zeigte, dass das gewählte Verfahren praxistauglich war. Wir legten das Ergebnis über die AG Sacherschließung offen und baten um Stellungnahmen. Die eingegangenen Äußerungen führten zu einigen Verbesserungen im Detail, nicht aber zu grundsätzlichen Änderungen.

Ein weiteres Fazit zog die Verbundzentrale: die Datenstruktur im SWB mit Titeldatei und verknüpfter Normdatei gewährleistet langfristig den aktuellen und gepflegten Stand der Schlagwortstammdatei. Änderungen aus der SWD werden zentral, laufend und automatisiert eingespielt. Demgegenüber führt das Datenmodell im BVB zu einem scheinbar höheren redaktionellen Aufwand, da eine Änderung an der Ansetzung eines SWD-Satzes in allen Titelsätzen, in denen die Ansetzung verwendet wurde, im Wesentlichen manuell nachvollzogen werden muss. Dies zeigte sich deutlich daran, dass ein recht großer Anteil der Schlagworte aus dem BVB, die nicht der SWD zugeordnet werden konnten, auf nicht nachvollzogene Änderungen in der SWD zurückzuführen sind. Als Beispiel sei genannt die nachträgliche Einführung eines Homonymenzusatzes bei einem Autor literarischer Werke: nicht nur die als Schlagwort verwendete Autorenansetzung, sondern alle verbundenen Sachtitelansetzungen sind zu überarbeiten. Die Verbundzentrale des SWB wird – auch aufgrund dieser Erfahrung – bei der Definition zukünftiger Datenmodelle deshalb besonderes Augenmerk darauf richten, diese bewährte Struktur nicht nur aufrecht zu erhalten, sondern nach Kräften zu verbessern.

Als Ergebnis des gesamten Projektes ergibt sich heute (Januar 2003) nach Abschluß der Einspielungen Folgendes: In 968.000 Titeln wurden 1.684.000 neue Schlagwortketten mit insgesamt 4.346.000 Schlagwortsätzen abgelegt. In gesamt 1.506.000 Titelaufnahmen wurden 2.207.000 Notationen der RVK mit ca. 202.000 unterschiedlichen Systemstellen verknüpft. Bedauert wird, dass die angelegten bzw. verwendeten RVK-Sätze außer der Notation und der Kennzeichnung

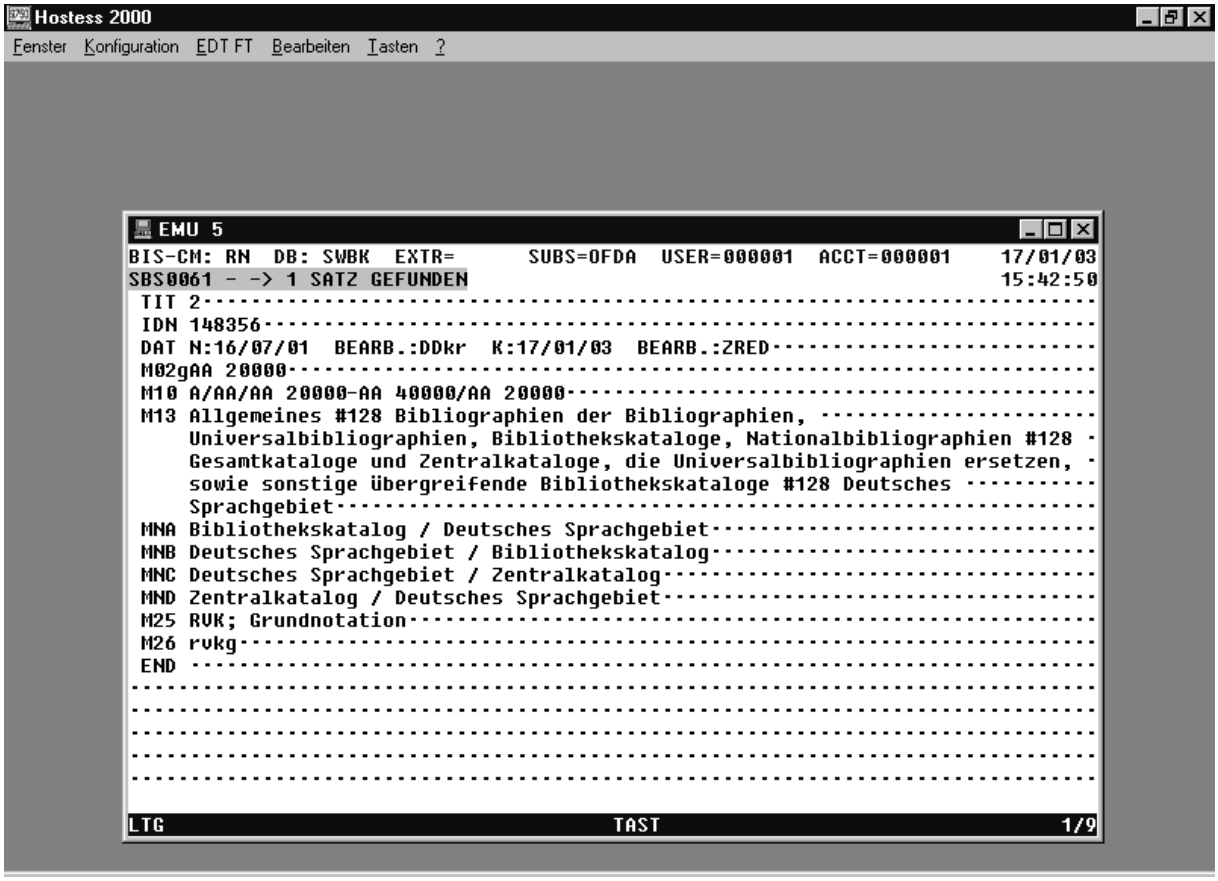
als RVK-Satz bislang eigentlich keine weiteren Informationen enthalten (z.B. Kennzeichnung als Grund- bzw. geschlüsselte Notation, Benennungen, Registereinträge, hierarchische Einordnung).

Den Nutzen für Recherche und Auskunft der zusätzlichen Sacherschließungsinformation an den insgesamt bearbeiteten 1.780.000 Titeln wird sicher die Zukunft zeigen.

Die Einspielung von Benennungen und Registereinträgen der RVK...
Schon 1994 hält die Veröffentlichung des DBI "Sacherschließung im OPAC" fest, dass die Evidenz der Systemstelle für den Bibliothekskunden nur durch die Wiedergabe der Benennung, die Suchbarkeit durch die Haltung der Registereinträge verwirklichbar ist. Seit den Anfangstagen des SWB steht sowohl in der regionalen als auch in der lokalen Notationsstammdatei ein Format zur Verfügung, das für diesen Zweck wenigstens als in vorläufiger Form geeignet erscheint, aber eigentlich nie ausgeschöpft wurde. Der Gedanke lag nahe, in Rücksprache mit der UB Regensburg und Zusammenarbeit mit der SLUB Dresden zu überlegen, ob ein geeignetes Verfahren und Programm entwickelt werden kann, wie Benennungen, Registereinträge, hierarchische Einordnung der einzelnen Systemstelle und gewisse redaktionelle Bemerkungen im SWB gehalten, recherchierbar gemacht und ausgeliefert werden können. Der Gedanke war insofern von besonderem Reiz, da die Registereinträge der RVK an der SWD abgeglichen sind (aus verschiedenen Gründen ist eine konsequente technische Verknüpfung nicht ratsam). Die Überlegungen zeigten, dass dieses Vorhaben durchführbar ist und die Daten von der UB Regensburg in verarbeitbarer Form zur Verfügung gestellt werden können.

Festgelegt wurde,
- dass die Kategorie M10 für die Wiedergabe der hierarchisch übergeordneten Systemstellen verwendet werden soll;
- dass die Kategorie M13 für die Wiedergabe der Benennungen der betroffenen samt der ihr übergeordneten Systemstellen verwendet werden soll;
- dass die Kategorien MNA-MNS für die Wiedergabe der SWD-gerechten Registereinträge benutzt werden sollen;
- dass Kategorie M25 für redaktionelle Hinweise genutzt wird.

Ein angereicherter Datensatz der RVK wird etwa dieses Aussehen haben:



SCREENSHOT AA 20000 – Deutsches Sprachgebiet

Derzeit (Januar 2003) laufen die letzten Absprachen zur endgültigen Gestalt dieser Sätze. Die Aufwertung der Sätze kann aber aus Kapazitätsgründen erst nach Abschluss der Übernahme der bayerischen Titelerschließung realisiert werden. Ihre Auslieferung im Rahmen der üblichen Datendienste des BSZ ist gewährleistet.

Der Versuch, nur die Grundnotationen der RVK mit diesen Informationen anzureichern, musste aus Gründen der RVK-inhärenten Datenstruktur aufgegeben werden. Eingespielt werden jetzt also die Benennungen, Registereinträge etc. sämtlicher möglichen Grund- und Schlüsselnotationen der RVK. Falls ein Satz noch nicht besteht, wird er im Rahmen dieses Projektes angelegt. Dies bietet den Vorteil, dass - werden die Benennungen eingespielt sein –nur noch Ergänzungen bzw. Änderungen der bestehenden Systematik nachvollzogen werden müssen. Die sonst zu erwartende beträchtliche Zahl laufend auftauchender neuer geschlüsselter Notationen wird nicht zu bearbeiten oder zu redigieren sein. Dieses Verfahren bietet aber auch eine Möglichkeit, im Rahmen der Einspielung der bayerischen sachlichen Titelerschließung entstandene RVK-Sätze einer ersten automatisierten Überprüfung zu unterziehen.

... führt zu einer erweiterten Verwendung natürlichsprachlicher Mittel in der Sacherschließung

Gemeinsame Überzeugung in allen Studien zur OPAC-Nutzung ist, dass der verbale sachliche Sucheinstieg die von der Nutzerschaft präferierte Recherchestrategie darstellt. Eine Gestaltung der Recherche, die die unterschiedlichen Erschließungsmethoden zu ihrem je eigenen Recht kommen lässt, wird am Ende aber nicht vernachlässigen, eine Möglichkeit anzubieten, die die verschiedenen Verfahren verbindet.

Neben RSWK/SWD und RVK zeichnet sich ab, dass aus mit der Dewey Decimal Classification erschlossenem Material zusätzliche Informationen gewonnen werden können. Bislang ist das Projekt der Übersetzung und des Aufbaus einer Normdatei DDC Deutsch nicht wirklich über das Vorbereitungs- und Verhandlungsstadium mit OCLC/ForrestPress als Lizenzgeber hinausgekommen. Das BSZ muss sich deshalb auf Begleitung bzw. Mitwirkung in DDC-Konsortium und Expertengruppe DDC beschränken, wird aber versuchen, die Projektergebnisse schnell einer breiten Verwendung zuzuführen, zumal da beabsichtigt ist, die Notationen der DDC mit ihren Registereinträgen an der SWD abzugleichen, wie dies für die RVK der Fall ist.

Prof. Albert Raffelt äußert in einer Mail an die Verbundzentrale diese Ansicht: "RVK ist für mich durchaus in Sacherschließungsperspektive interessant. Eben deshalb kam meine Frage, ob man nicht die gesamten Sacherschließungsdaten (RVK, RSWK, sonstige Fremddaten) irgendwann in einem Erschließungssystem à la maniere d'Osiris zusammenführen kann, das dann zusätzlich zu der sehr präzisen Suche in Systemen wie RSKW/SWD einen weiteren Einstieg bieten würde." Sicher weist diese Äußerung noch in die Zukunft; dennoch wird die Verbindung verbesserter Arbeitswerkzeuge im Geschäftsgang zu einer schnelleren Erschließung beitragen, diese wiederum in Verbindung mit eventuell möglichen Projekten der zusätzlichen Fremdleistungsübernahme zu einer weiteren Steigerung der Erschließungsdichte. Die Kombination verschiedener Erschließungsmethoden kann dann die Zugänglichkeit der angebotenen Dokumente über Expertensysteme und linguistisch bzw. wortstatisch unterstützte Retrievalsysteme verbessern. Verbindendes Glied wird sein die unterlegte Verwendung natürlichsprachlicher Mittel, besonders im Kontext der SWD, an deren Nutzung, Gestaltung und Ausbau die Teilnehmereinrichtungen mit der Verbundzentrale des SWB weiter festhalten werden.

Stefan Wolf
Bibliotheksservice-Zenrum Baden-Württemberg
Universität Konstanz
D-78457 Konstanz
Tel.: +49 (0) 75 31/88 30 50
E-Mail: stefan.wolf@bsz-bw.de