

Monika Kuberek

## **KOBV-Volltextserver**

—

## **Langfristiger Zugriff auf Kluwer-, Springer- und Elsevier-Zeitschriften**

**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**



# KOBV-Volltextserver – Langfristiger Zugriff auf Kluwer-, Springer- und Elsevier-Zeitschriften

Monika Kuberek

Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV)  
im  
Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)

ZIB Report 05-59 (November 2005)

## Abstract

Seit einigen Jahren lizenzieren Bibliotheken mit erheblichem finanziellen Aufwand elektronische Zeitschriften. Anders als bei einer Papierausgabe ist der dauerhafte Zugriff auf die bezahlten Dokumente allerdings nicht garantiert: Die e-Zeitschriften liegen auf dem Verlags-server, und der Verlag schaltet den Zugriff (meist IP-Range des Campus) auf seinem Server frei. Wird der Zugriff von Verlagsseite abgeschaltet, erlöschen sämtliche Zugriffsrechte, auch auf die in der Vergangenheit lizenzierten und bezahlten Zeitschriften. Auf die neuen Abonnementbedingungen hat das Friedrich-Althoff-Konsortium (FAK) reagiert und in seinen Verträgen den Erwerb der Archivdaten beim Auslaufen eines Vertrages festgeschrieben.

Im Rahmen eines Projektes baut die KOBV-Zentrale einen Zeitschriftenserver auf, um den Zugriff auf die lizenzierten elektronischen Zeitschriften auch nach Ablauf der Lizenzverträge zu gewährleisten. Den Grundstock bilden die vom FAK in den Jahren 1997-2003 lizenzierten elektronischen Zeitschriften der Verlage Kluwer Academic Press, Springer und Elsevier – ein Volumen von rund 1.600 Zeitschriftentiteln mit knapp 1.400.000 Artikeln.

Beim Aufbau des Zeitschriftenservers kommt der vertraglich-organisatorischen Komponente eine ebenso wichtige Rolle zu wie der technischen Realisierung. Hier hat die KOBV-Zentrale ein transparentes Verfahren konzipiert und umgesetzt, um für die Verlage die notwendige Vertrauensbasis zu schaffen und gleichzeitig den Einrichtungen ihren berechtigten Zugriff auf die Zeitschriften-Volltexte zu sichern.

Die Zeitschriftenartikel werden sowohl im Rahmen des KOBV-Volltextservers, einem neuen Internet-Dienst der KOBV-Zentrale, zugänglich gemacht – volltext-indiziert mit der Suchmaschine Swish-e – als auch integriert in die Metadatensuche und den Open-Linking-Dienst des KOBV-Portals. Während die Metadatenrecherche und die Sicht auf die Abstracts für alle offen sind, ist der Zugriff auf die Artikel-Volltexte auf die an den Verträgen beteiligten Einrichtungen beschränkt. Dazu wurde ein Authentifizierungsverfahren auf der Basis von IP-Ranges eingerichtet.

Der vorliegende Text ist die schriftliche Fassung eines gleichnamigen Vortrages auf der ASpB-Tagung 2005 "Spezialbibliotheken zwischen Auftrag und Ressourcen" der Arbeitsgemeinschaft der Spezialbibliotheken, die vom 06.-09. September 2005 in der Technischen Universität München stattfand.

**Keywords:** Archivierung, Authentifizierung, Autorisierung, Bibliotheksverbund, Dokumentenserver, elektronische Dokumente, elektronische Zeitschriften, KOBV, KOBV-Volltextserver, Kooperativer Bibliotheksverbund Berlin-Brandenburg, Langzeitverfügbarkeit, lizenzierte Zeitschriften, Suchmaschinenteknologie, Swish-e, Volltextindex, Volltexte, Volltextserver, Zeitschriftenserver, Zeitschriftenartikelservers, Verfügbarkeit

CR: H.0K, 4.0

## Inhalt

Vorbemerkung.....	3
Projekt Zeitschriftenserver der KOBV-Zentrale in Kooperation mit dem FAK.....	3
Organisatorische Komponente – Gewährleistung der Einhaltung der Zugriffsrechte .....	4
Technische Realisierung – Aufbau des KOBV-Zeitschriftenartikelservers.....	5
Der KOBV-Volltextserver – Volltextindex über Zeitschriftenartikel und elektronische Dokumente.....	6
Integration in die KOBV-Internet-Dienste und Ausblick.....	8

## Vorbemerkung

Seit einigen Jahren lizenzieren Bibliotheken mit erheblichem finanziellen Aufwand elektronische Zeitschriften. Anders als bei einer Papierausgabe ist der dauerhafte Zugriff auf die bezahlten Dokumente allerdings nicht garantiert: Die Bibliotheken erwerben in der Regel lediglich die Nutzungsrechte. Die elektronischen Zeitschriften werden nicht an die Bibliotheken ausgeliefert, sondern verbleiben physisch bei den Verlagen. Sie liegen auf dem Verlagsserver, und der Verlag schaltet den Zugriff (meist IP-Range des Campus) auf seinem Server frei. Wird der Zugriff von Verlagsseite abgeschaltet, erlöschen sämtliche Zugriffsrechte, auch auf die in der Vergangenheit lizenzierten und bezahlten Zeitschriften. Dies ist beispielsweise der Fall beim Auslaufen eines Lizenzvertrages. Aber auch andere, unvorhergesehene Gründe – wie Vertragsstreitigkeiten, die Kündigung einer Lizenz (sei es von Seiten des Verlages oder von Seiten des Lizenznehmers), das Einstellen einer Produktlinie, veränderte Bedingungen nach einer Verlagsübernahme, um nur einige solcher "widrigen Umstände" zu nennen – können dazu führen, dass die Bibliothek den Zugriff auf ihre lizenzierten Zeitschriften verliert. Auf diese Umstände und die neuen Abonnementbedingungen bei elektronischen Zeitschriften hat das Friedrich-Althoff-Konsortium (FAK)<sup>1</sup> reagiert: Die Verträge wurden so ausgehandelt, dass der Erwerb der Archivdaten beim Auslaufen eines Vertrages festgeschrieben ist, verbunden mit dem Erwerb der Zugriffsrechte für die berechtigten Bibliotheken und dem Recht, die Daten auf einem eigenen Server bzw. einem Server eigener Wahl aufzulegen.

## Projekt Zeitschriftenserver der KOBV-Zentrale in Kooperation mit dem FAK

War die Sicherung der Zugriffsrechte ein erster, wesentlicher Schritt zur langfristigen Verfügbarkeit der lizenzierten elektronischen Zeitschriften, so begann die eigentliche Herausforderung mit dem Auslaufen der ersten Verträge im Jahre 2003. Galt es doch, die Daten zu speichern und den Zugriff auf die Zeitschriftenartikel zu realisieren, wobei gewährleistet werden musste, dass nur berechtigte Nutzer einen Zugriff erhalten. Kluwer Academic Press war der erste Verlag, von dem Datenlieferungen eintrafen, gefolgt von Springer und Elsevier. Im Rahmen eines Projektes hat die KOBV-Zentrale – in Kooperation mit dem FAK – einen Server für die Archivdaten der lizenzierten elektronischen Zeitschriften aufgebaut. Zielsetzung ist es, den Zugriff auf die Volltexte der Zeitschriftenartikel langfristig zu gewährleisten.

Das Projekt der KOBV-Zentrale ist im Zusammenhang mit dem Projekt "Verteilter Dokumentenserver" (VDS) zu sehen, das die Arbeitsgemeinschaft der Verbundsysteme auf Initiative des KOBV im Jahre 2003 gestartet hat.<sup>2</sup>

Den Grundstock für den KOBV-Zeitschriftenartikelservers bilden die vom FAK in den Jahren 1997-2003 lizenzierten elektronischen Zeitschriften der Verlage Kluwer Academic Press,

---

<sup>1</sup> Das Friedrich-Althoff-Konsortium (FAK) ist das regionale Konsortium für Berlin und Brandenburg, in dem sich die wissenschaftlichen Einrichtungen der Länder Berlin und Brandenburg zusammen geschlossen haben, um ihre Nutzerinnen und Nutzer mit wissenschaftlicher Information aus elektronischen Veröffentlichungen zu versorgen. Die Geschäftsstelle des FAK ist in der KOBV-Zentrale angesiedelt. Weitere Informationen zum FAK: <http://www.fak.de> .

<sup>2</sup> Das Vorhaben der Verbünde wurde auf dem Bibliothekskongress 2003 vorgestellt und ist beschrieben in: Joachim Lügger: Zusammenarbeit mit vascoda im Rahmen des Verteilten Dokumentenservers (VDS). – In: Information macht Bildung. – Hrsg. von Georg Ruppelt u. Gabriele Beger. Wiesbaden: Dinges & Frick 2004. S. 98-116. –  
Volltext unter: <http://www.zib.de/Publications/Reports/ZR-04-30.pdf> .

Springer und Elsevier. Insgesamt handelt es sich um ein Volumen von rund 1.600 Zeitschriftentiteln mit knapp 1.400.000 Artikeln (Aufsätze, Rezensionen usw.). An den Verträgen waren – in unterschiedlichen Konstellationen – 39 verschiedene Einrichtungen aus 9 Bundesländern beteiligt.

- *Kluwer Academic Press (jetzt Springer)*
  - 775 Zeitschriften
  - 237.000 Artikel
  - Vertragslaufzeit: 1997-2003
  - Lizenznehmer waren 39 Einrichtungen aus 9 Bundesländern (Berlin, Brandenburg, Bremen, Hamburg, Mecklenburg-Vorpommern, Niedersachsen, Sachsen, Sachsen-Anhalt, Thüringen)
- *Springer*
  - 211 Zeitschriften
  - 131.000 Artikel
  - Vertragslaufzeit: 1998-2003
  - Lizenznehmer waren 9 Einrichtungen aus 2 Bundesländern (Berlin, Brandenburg)
- *Elsevier*
  - 618 Zeitschriften
  - 1.009.000 Artikel
  - Vertragslaufzeit: 1997-2003
  - Lizenznehmer waren 8 Einrichtungen aus 2 Bundesländern (Berlin, Brandenburg)

### **Organisatorische Komponente – Gewährleistung der Einhaltung der Zugriffsrechte**

Beim Aufbau des Zeitschriftenservers kam der vertraglich-organisatorischen Komponente eine ebenso wichtige Rolle zu wie der technischen Realisierung. Es galt, mit den Verlagen Absprachen zu treffen und ein Verfahren zu entwickeln, das verlässlich ist und bei den Verlagen die notwendige Vertrauensbasis dafür schafft, dass die Einhaltung der Zugriffsrechte auf dem verlagsfremden Server gewährleistet ist: Auch für die Zukunft muss sicher gestellt sein, dass nur die Nutzer der Einrichtungen, die die Zeitschriften in der Vergangenheit lizenziert hatten, die Volltexte der Artikel einsehen können. Gleichzeitig galt es, den Einrichtungen ihren berechtigten Zugriff auf die Artikel-Volltexte zu sichern.

Die Zugriffsrechte jeder berechtigten Einrichtung sind anhand folgender Komponenten zu bestimmen: IP-Range, Zeitraum der Lizenznahme und lizenzierte Zeitschriftentitel. Auf der Basis dieser Komponenten wurde das technische Verfahren der Zugriffskontrolle implementiert (s.u.). Zur Verifizierung der Angaben hat die KOBV-Zentrale parallel zur Implementierung im Zusammenspiel der Beteiligten – dem FAK als vertragschließender Instanz, den Bibliotheken als Lizenznehmern, den Verlagen als Lizenzgebern und der KOBV-Zentrale als "technischer Schaltstelle" – ein transparentes Verfahren konzipiert und umgesetzt. Da es sich um Archivdaten handelt, sind Zeitraum der Lizenznahme und lizenzierte Zeitschriftentitel konstant; lediglich der IP-Range ist variabel. Auch für die nachträgliche Änderung einer IP-Adresse ist ein Verfahren mit den Verlagen vereinbart, das ebenfalls auf eine hohe Trans-

parenz setzt. Nachdem die Verlage ihre Zustimmung zu dem Verfahren gegeben hatten, wurde der Zugriff auf die Zeitschriftenartikel im August 2005 für die Nutzer freigeschaltet.<sup>3</sup>

## **Technische Realisierung – Aufbau des KOBV-Zeitschriftenartikelservers**

Anhand der Archivdaten von Kluwer Academic Press realisierte die KOBV-Zentrale den ersten Prototypen des KOBV-Zeitschriftenartikelservers. Die Kluwer-Datenlieferung umfasste – ebenso wie die von Springer und Elsevier – die Artikel-Volltexte und die Metadaten. Während die Volltexte als PDF-Dateien<sup>4</sup> geliefert wurden, einem Standardformat, das sich inzwischen bei den Verlagen als Speicherformat etabliert hat, wiesen die Metadaten ganz unterschiedliche, verlagseigene Formate auf. Mit erheblichem Aufwand wurden die Metadaten analysiert und in eine XML-Struktur überführt. Weiterhin waren grundlegende Problematiken im Zusammenhang mit den Artikel-Volltexten zu lösen – angefangen vom enormen Datenvolumen, über fehlende Verifikation (Sind die gelieferten Dateien tatsächlich identisch mit den Verlagsdaten?) und Synchronisation (Wurden alle Artikel geliefert?) bis hin zum Handling der diversen PDF-Versionen (allein in der Kluwer-Lieferung mehr als 100 PDF-Versionen).<sup>5</sup>

Die Artikel-Volltexte und die Metadaten sind in einem Dateisystem gespeichert. Die Daten liegen auf einem Server in der KOBV-Zentrale, die – vor dem Hintergrund der gut ausgebauten technischen Infrastruktur des ZIB – zuverlässige Datensicherungs- und Backup-Dienste sowie die langfristige Archivierung der Daten gewährleisten kann. Was den langfristigen Zugriff angeht, hat die KOBV-Zentrale – insbesondere aufgrund der Personalkapazitäten, die die aktive Beteiligung an einschlägigen Entwicklungsprojekten wie beispielsweise Nestor<sup>6</sup> nicht zulassen – einen pragmatischen Weg eingeschlagen, indem sie die Volltexte als PDF-Dokumente speichert. PDF ist als Standard-Speicherformat auf absehbare Zeit weitflächig im Einsatz. Die KOBV-Zentrale verfolgt sehr genau die Entwicklungen und wird – sobald es opportun ist – die PDF-Dokumente auf ein sich in der Zukunft etablierendes Speicherformat migrieren.

Um die Zeitschriftenartikel suchbar zu machen, hat die KOBV-Zentrale Metadaten und Volltexte mittels Suchmaschinenteknologie indiziert. Im Einsatz ist derzeit der Suchmaschine swish-e, ein Open-Source-Produkt. Auf diese Weise wurde jeweils ein Volltextindex der Kluwer-Daten und der Springer-Daten aufgebaut.<sup>7</sup> Swish-e erlaubt die virtuelle Zusammenführung verschiedener Indices, um sie mit einer einzigen Suche zu durchsuchen. Diese Eigenschaft hat sich die KOBV-Zentrale zunutze gemacht und die Indices im Hintergrund "zusammen geschaltet". Die Ergebnisse werden bei der Anzeige dynamisch gemergt und einem Ranking unterzogen. Für den Nutzer sieht es aus, als suche er in einem einzigen Datenbestand.

Für die Authentifizierung und Autorisierung wurde ein IP-kontrolliertes Verfahren implementiert, das mit dem FAK, den Bibliotheken und den Verlagen abgestimmt ist. Jeder Nutzer

<sup>3</sup> Der KOBV-Volltextserver ist erreichbar unter: <http://digibib.kobv.de> .

<sup>4</sup> PDF = Portable Document Format

<sup>5</sup> Die grundlegenden Problematiken wurden anhand der Kluwer-Daten analysiert und aufgezeigt; s. Thorsten Koch: Verteilter Dokumentenspeicher – Erfahrungen mit den Kluwer-Daten des Friedrich-Althoff-Konsortiums. – ZIB-Report 03-50 (Dezember 2003). – Volltext unter: <http://www.zib.de/Publications/Reports/ZR-03-50.pdf> .

<sup>6</sup> nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen in Deutschland: <http://www.langzeitarchivierung.de>

<sup>7</sup> Die Elsevier-Daten werden derzeit indiziert und noch in 2005 ebenfalls zur Verfügung gestellt.

kann nach allen Artikeln suchen und die Kurzfassung (Abstract) der Aufsätze einsehen. Um den Volltext eines Artikels zu sehen, muss er sich allerdings auf dem Campus einer Einrichtung befinden, die diese Zeitschrift in den entsprechenden Jahren abonniert hatte. Bei den Einrichtungen handelt es sich, wie oben erwähnt, um Hochschulen oder wissenschaftliche Institute – vorwiegend mit Sitz in Berlin und Brandenburg, teilweise bundesweit.

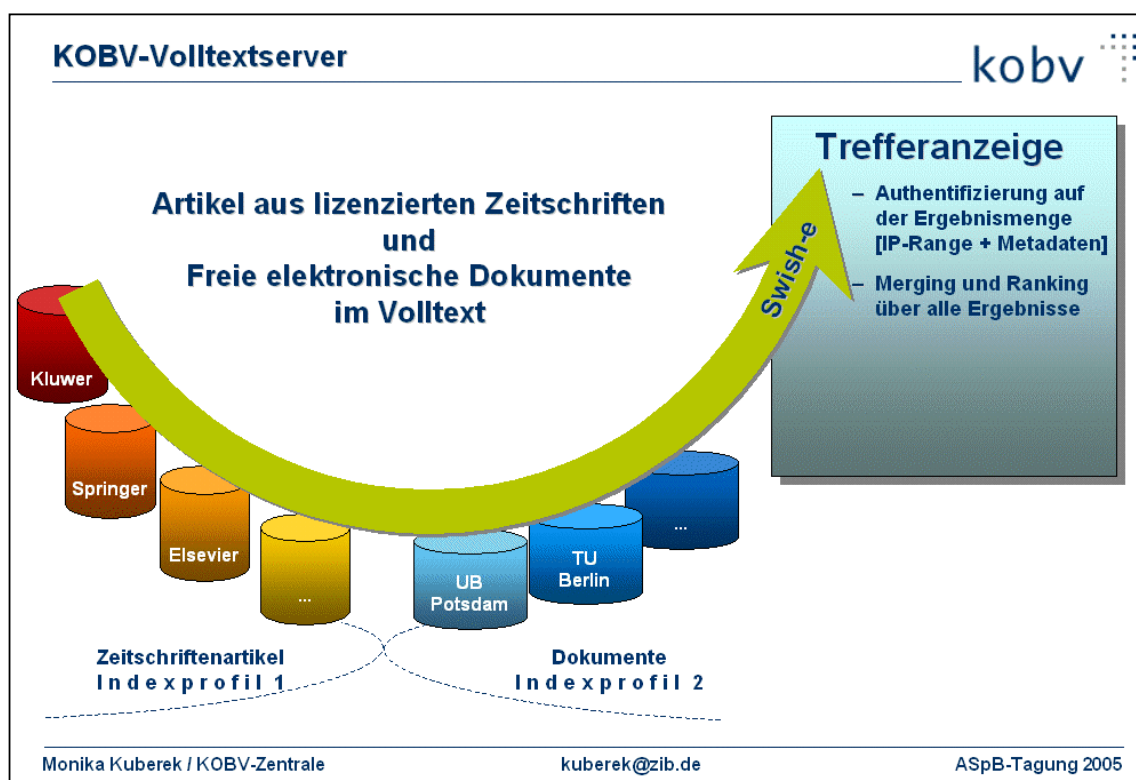


Abb. 1: KOBV-Volltextserver: Zugriff auf Artikel aus lizenzierten Zeitschriften und auf freie elektronische Dokumente im Volltext

## Der KOBV-Volltextserver – Volltextindex über Zeitschriftenartikel und elektronische Dokumente

Die Eigenschaft von swish-e, die Volltextsuche über verschiedene Indices durchzuführen sowie unterschiedliche Indexp Profile anzulegen und zusammen zu führen, hat sich die KOBV-Zentrale in einer weiteren Hinsicht zunutze gemacht und den KOBV-Volltextserver aufgebaut, der einen virtuellen Volltextindex über die lizenzierten Zeitschriftenartikel und freie elektronische Dokumente<sup>8</sup> realisiert. Mit einer einzigen Suche kann der Nutzer in *allen Volltexten* auf dem KOBV-Server recherchieren (s. Abb. 1).

<sup>8</sup> Dabei handelt es sich um die elektronischen Dokumente der regionalen Dokumentenserver, die von der KOBV-Zentrale als Serviceleistung in einem Konsortialmodell betrieben werden. Die einzelnen Dokumentenserver sind ebenfalls mit swish-e Volltext-indiziert. Die Volltextindices können lokal einzeln abgesucht werden und werden gleichzeitig für die regionale Suche in der KOBV-Zentrale virtuell "zusammen geschaltet".



Aufbauend auf dem KOBV-Zeitschriftenartikelservers ist die KOBV-Zentrale damit einen Schritt weiter gegangen und hat mit dem KOBV-Volltextserver eine oft diskutierte Vision realisiert – nämlich die Trennung zwischen Zeitschriftenartikeln und Dokumenten aufzuheben und unterschiedliche Materialien in einem gemeinsamen Volltextindex suchbar zu machen.

Für den Nutzer, der "seinen" Text möglichst sofort auf dem Bildschirm haben möchte, ist es irrelevant, ob es sich bei dem gesuchten Volltext um einen Zeitschriftenartikel oder ein elektronisches Dokument handelt. Auch die Tatsache, dass die Zeitschriftenartikel lizenziert sind, ist für die Suche zunächst ohne Belang. Erst beim Zugriff auf den Volltext kommt dieses Moment zum Tragen: Nur der Nutzer, der sich auf dem Campus einer berechtigten Einrichtung befindet, erhält den Artikel im Volltext. Aber auch wenn er keine Berechtigung hat, kann er zumindest das Abstract einsehen und erhält die Information über die berechtigten Einrichtungen, auf deren Campus er den Volltext erhalten kann.

Durch ein einfaches Ampelschema ist für den Nutzer auf den ersten Blick transparent, ob er den Zugriff auf den Volltext hat oder nicht. Während für die frei zugänglichen elektronischen Dokumente die Ampelschaltung immer auf grün steht, waren für die lizenzierten Zeitschriften die Zugriffsrechte abzubilden. Dazu hat die KOBV-Zentrale die Anzeige der Zugriffsrechte dynamisch implementiert – abhängig vom Zugriffsort des Nutzers. Befindet er sich auf dem Campus einer berechtigten Einrichtung, signalisiert ihm eine grüne Ampel, dass er den Zugriff auf den Volltext hat. Ansonsten zeigt ihm eine rote Ampel an, dass er keine Zugriffsberechtigung besitzt.



Abb. 2: KOBV-Volltextserver – Trefferanzeige: Lizenzierte Zeitschrift (oben; hier mit roter Ampel) und freies elektronisches Dokument (unten; mit grüner Ampel).

## Integration in die KOBV-Internet-Dienste und Ausblick

Hat die KOBV-Zentrale mit KOBV-Volltextserver einen neuen Internet-Dienst geschaffen, so hat sie – als weiteren Sucheinstieg für die Nutzer – parallel dazu den Zeitschriftenartikelservers in die bestehenden Internet-Dienste integriert. Als Ressource wurde der Server in das KOBV-Portal eingebunden und kann dort über den Fachgebiets-Einstieg gefunden werden.<sup>9</sup> Dazu wurde der Server – entsprechend der fachlichen Ausrichtung der Zeitschriftenartikel – verschiedenen Fachgebieten zugeordnet (u.a. Naturwissenschaften, Technik, Medizin und Gesundheit, Recht usw.). Der Volltextindex ist auch im KOBV-Portal durchsuchbar und kann – zusammen mit anderen durchsuchbaren Ressourcen – in einer parallelen Suche durchsucht werden. Die Zugriffsrechte werden über SFX verwaltet, das als Reference-Linking-Werkzeug im KOBV-Portal genutzt wird. Befindet sich ein Nutzer auf dem Campus einer berechtigten Einrichtung, so kann er den Volltext über das SFX-Menü abrufen.

Wie sehen die nächsten Schritte im Projekt Zeitschriftenserver aus? Sobald – noch in 2005 – die Elsevier-Zeitschriften freigegeben sind, hat die KOBV-Zentrale zunächst einmal alle Archivdaten aus ausgelaufenen FAK-Verträgen für die Nutzung zugänglich gemacht. In der Planung ist die Implementierung eines automatisierten Verfahrens zwischen den Bibliotheken und den Verlagen zur Aktualisierung von IP-Adressen, um den Workflow zu optimieren. Die Realisierung dieser Maßnahme wird – wie schon die bisherigen Schritte im Zusammenhang mit den Zugriffsrechten – in enger Kooperation mit allen Beteiligten erfolgen, um den Verlagen auch für die Zukunft die notwendige Vertrauensbasis zu erhalten. Weiterhin steht die Konzeption eines langfristigen Finanzierungsmodelles für den KOBV-Zeitschriftenartikelservers an – dies auch vor dem Hintergrund, dass FAK und KOBV nicht identisch sind. Schließlich wird die KOBV-Zentrale das Konzept des langfristigen Zugriffes auf lizenzierte Zeitschriften in Zusammenarbeit mit der Arbeitsgemeinschaft der Verbundsysteme im Rahmen des Projektes "Verteilter Dokumentenserver (VDS)" weiter vorantreiben.

---

<sup>9</sup> Das KOBV-Portal – Digitale Bibliothek Berlin-Brandenburg ist erreichbar unter:  
<http://digibib.kobv.de> :