

Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

JOACHIM LÜGGER

Vorhaben
Offene Navigation und Suchmaschinen
in Verbänden, Konsortien und den Wissenschaften
Schlussbericht zum
Teilprojekt des BMBF-Vorhabens
Kompetenznetzwerk Neue Dienste, Standardisierung, Metadaten

Das Vorhaben wurde unter dem Kennzeichen 08C5929 vom BMBF gefördert.

Abstract:

Die harmonische Integration der Navigation und Suche in lizenzierten Journalen und gleichzeitig in freien digitalen Dokumenten unter einer einheitlichen konsistenten Nutzeroberfläche ist eines der ungelösten F&E-Probleme der Fachinformation. Hierfür sollen Elemente des Invisible Web und des Visible Web unter Berücksichtigung offener Standards nahtlos miteinander verbunden werden. Dem Projekt liegt ein Modell mit Internet-Index, Metasuche und Open Linking über verteilten heterogenen Speichern zu Grunde: Verschiedenste Server, digitale Referenzen in Artikeln und Dokumenten, Links in Datenbanken und auf Bestelldienste sollen unter Berücksichtigung von Standort-, Studien- und Lernbedingungen kooperativ miteinander vernetzt werden. Die Leistungsfähigkeit des Modells soll in Pilotimplementierungen getestet und für eine breite Anwendung vorbereitet werden. Auf dieser Basis soll das Vorhaben Verteilter Zeitschriftenserver der AG der Verbundsysteme in eigenen Teilprojekten kooperativ initiiert werden, das jetzt in das Vorhaben Verteilter Dokumentenserver von vascoda integriert ist.

Keywords:

Metasuchmaschine, Internet-Suchmaschinen, kontextsensitives Linking, OpenURL, Verteilter Zeitschriftenserver, Verteilter Dokumentenserver, Virtueller Dokumentenspeicher, Portalsystem, Nahtlose Navigation vascoda

1. Formalangaben

1.1 Förderkennzeichen:	08C5929
2.2 Thema des Vorhabens:	Offene Navigation und Suchmaschinen in Verbänden, Konsortien und den Wissenschaften
1.3 Zuwendungsempfänger:	Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)
1.4 Ausführende Stelle:	Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)
1.5 Projektleiter:	Prof. Dr. M. Grötschel
1.6 Bewilligungszeitraum:	01.10.02 bis 31.03.03
1.7 Berichtszeitraum:	01.02.02 bis 31.03.03

2. Wissenschaftlich-technische Ziele und Vorgehensweise

2.1 Darstellung des erweiterten Projektrahmens

Angesichts der Kürze des Projektzeitraums und der besonderen Art des Projektes haben wir uns dazu entschieden, den Schlussbericht um eine „Vor- und Nachgeschichte“ zu erweitern und in freier Form zu gestalten, um in ihm auch die Beteiligung eines Kreises von Personen, die an den Diskussionen um den Verteilten Zeitschriftenserver (VZS¹) bzw. Verteilten Dokumentenserver (VDS) in vascoda [1] beteiligt waren, zum Ausdruck bringen zu können. Der Schlussbericht selbst konzentriert sich dabei auf die Darstellung der Projektergebnisse zum Ende des Bewilligungsraumes, also auf eine Zeit, in der die Diskussionen um ein Zusammengehen der Verbände mit vascoda erst begannen.

Der Schlussbericht geht aber nicht nur auf die Vorgeschichte ein, die nicht eigentlich der Gegenstand des hier vorliegenden Berichtes ist, sondern auch auf Ereignisse und Ergebnisse, die erst im Anschluss an das Vorhaben eintraten bzw. erzielt werden konnten. Diese wurden in erster Linie in enger Kooperation mit und von den Partnern der „Arbeitsgemeinschaft der Verbundsysteme“ (AGVerbund) erreicht. Das ZIB und der in ihm beheimatete KOBV haben das Projekt nach seinem formalen Abschluss am 31.03.03 – wegen seiner grundsätzlichen Bedeutung – aus eigener Kraft, im eigenen Interesse und in enger Kooperation mit der AGVerbund und ausgewählten Entwicklungszentren weitergeführt. Die Arbeiten, Projektdefinitionen und Berichte aus diesen Aktivitäten reichen bis an das Jahresende 2003 heran.

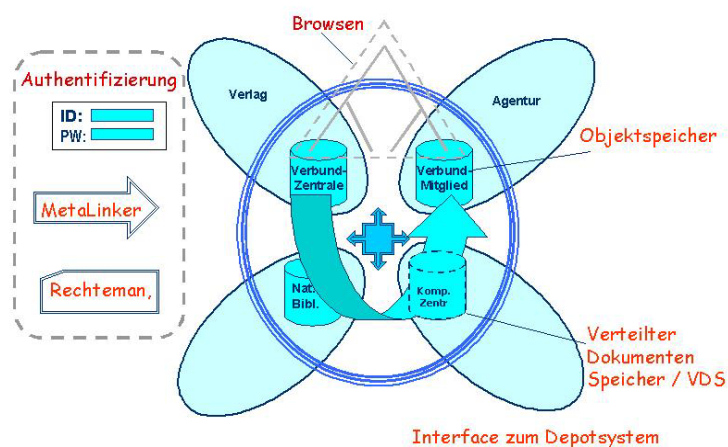
In diesem Zusammenhang ist auch der „Rahmenantrag“ des Vorhabens „Aufbau eines ‚Verteilten Dokumentenservers‘ (VDS) zum Aufbau einer offenen digitalen Bibliothek für die Wissenschaft“ [2] zu sehen, dessen technisch-organisatorischer Kern vom ZIB/KOBV und seinen Partnern im Kontext des Zusammengehens der AGVerbund mit vascoda erstellt wurde.

Zum Rahmenantrag und den vorbereitenden Projektdefinitionen und Abstimmungen hat auch das hier vorliegende Vorhaben einen wesentlichen Beitrag geleistet. Die Koordinationsaktivi-

¹ Wir verwenden im Nachfolgenden die Bezeichnungen „VDS des KOBV“ für den „Verteilten“ oder „Virtuellen Dokumentenspeicher“ des KOBV, VZS für das Projekt „Verteilter Zeitschriftenserver“ der AGVerbund und VDS für dessen Verallgemeinerung zum „Verteilten Dokumentenserver“ im Kontext von vascoda. Alle Vorhaben verwenden dieselbe generische Modellvorstellung, sind aber in Bezug auf den Content unterschiedlich ausgerichtet. Der VZS ist, so gesehen, im VDS enthalten. Zeitschriftenartikel sind nur ein besonderer Materialfall.

täten des ZIB/KOBV wären nicht ohne eine ständige technische Überprüfung und Verifikation und wohl auch nicht ohne die aktuellen Arbeiten des KOBV am KOBV-Informationsportal möglich gewesen. Die wesentlichen technischen Grundlagen für die Überprüfung der Konzepte lieferte der Aufbau eines substantiellen Speichers digitaler Zeitschriften (ca. 250.000 Artikel mit zugehörigen Metadaten aus Kluwer-Zeitschriften) [3] und deren Indexierung (ein Volltextindex und ein Metadaten-basierter Index) mittels der Suchmaschine FAST der Firma Fast Search & Transfer² [4]. Die Erfahrungen des KOBV beim Einsatz des Portal- und Open-Linking-Systems MetaLib/SFX sind ebenfalls zu diesen Basis-Erfahrungen zu zählen.

Bevor sich der Schlussbericht nun dem engeren Thema dieses Projektes zuwendet, soll – zur Einführung und zur Wahrung der Perspektive – eine knappe und aktuelle technische Zusammenstellung der VDS-Projekte³ und ihrer Phasen im Kontext des VDS-Gesamtkonzeptes der AGVerbund in vascoda gegeben werden.



Abbild 1. Modellvorstellung des Verteilten Zeitschriftenservers und des VDS

I. Spiegel und Speicher, Metadaten und Ressourcen

- | | |
|---------------------------|---|
| a. Verbünde, Bibliotheken | Spiegel- und Speicher-Server (Volltextserver) |
| b. Bibliotheken, Verbünde | Einbindung in lokale Portale (incl. Open Linking) |
| c. BSZ und Verbünde *) | Metadaten-Verwaltung |
| d. KOBV *) | Virtuelle Dokumentenspeicher |

II. Verteilte Zugriffskontrolle und Open Linking

- | | |
|-------------------------|---|
| a. UB-Freiburg/ReDI *) | Verteilte Authentifizierung / Autorisierung |
| b. UB-Regensburg/EZB *) | Lizenzkontrolle / Rechtmanagement |

III. Nahtlose Navigation und Suche

- | | |
|-------------------------|------------------------------------|
| a. UB-Bielefeld, HBZ *) | Erprobung von Internet-Technologie |
|-------------------------|------------------------------------|

IV. Verteilter Betrieb, Dauerhaftigkeit

- | | |
|--------------------------|---|
| a. Verbünde/Bibliotheken | Verteilter Betrieb |
| b. ZIB/KOBV | Internet-Tools für Zusammenspiel VDS-Module |

In Vorbereitung bzw. assoziiert sind die folgenden Projekte:

- | | |
|------------|--|
| a. BSB/BVB | Nahtlose Navigation und Open Linking (lokal) |
|------------|--|

² Die Firma FAST wurde inzwischen von der Firma Overture übernommen.

³ Projekte, zu denen BMBF- oder DFG-Förderanträge vorliegen, sind mit einem Sternchen gekennzeichnet

- | | |
|------------|--|
| b. SBB/ZDB | Lokalisierung lokaler Print-Materialien |
| c. BSB | Überregionale Koordination von Beschaffungsmodellen und Angebotsformen |
| d. DDB | Interface Depotsystem ~ Spiegel/Speicher |

2.2 Darstellung der wissenschaftlich-technischen Ziele

Das Teilprojekt „Offene Navigation und Suchmaschinen“, das als Teil des Vorhabens „Kompetenznetzwerk Neue Dienste, Standardisierung, Metadaten“ bewilligt wurde, verstand sich, etwas verkürzt ausgedrückt, in erster Linie als Vorprojekt zu dem damals noch in Vorbereitung befindlichen Vorhaben „Verteilter Zeitschriftenserver“ (VZS) der AGVerbund. Hierzu ein Auszug aus der Kurzfassung der Vorhabenbeschreibung aus dem Antrag für ein BMBF-Projekt „Verteilter Contentserver“⁴:

„Die harmonische Integration der Navigation und Suche in lizenzierten Journalen und gleichzeitig in freien digitalen Dokumenten unter einer einheitlichen konsistenten Nutzeroberfläche ist eines der ungelösten F&E-Probleme der Fachinformation. Hierfür sollen Elemente des Invisible Web und des Visible Web unter Berücksichtigung offener Standards nahtlos miteinander verbunden werden.

Dem Projekt liegt ein Modell mit Internet-Index, Metasuche und Open Linking über verteilten heterogenen Speichern zu Grunde: Verschiedenste Server, digitale Referenzen in Artikeln und Dokumenten, Links in Datenbanken und auf Bestelldienste sollen unter Berücksichtigung von Standort-, Studien- und Lernbedingungen kooperativ miteinander vernetzt werden. Die Leistungsfähigkeit des Modells soll in Pilotimplementierungen getestet und für eine breite Anwendung vorbereitet werden.

Auf dieser Basis soll das Vorhaben Verteilter Zeitschriften-Server der AG der Verbundsysteme und der Konsortien in eigenen Teilprojekten kooperativ initiiert werden. Das Modell soll auch, beispielhaft für die Wissenschaften, im Vorhaben World Mathematics Library der IMU Anwendung finden.“

Durch die Annahme der Bewilligungsaufgaben hatte sich ZIB/KOBV dazu verpflichtet, die

- Chancen und Risiken der vorgeschlagenen Maßnahmen eingehend zu analysieren,
- Verwertung der Ergebnisse sicherzustellen (insbesondere bzgl. der Vereinbarung der Aufgabenteilung und der Rechte und Pflichten der Kooperationspartner).

Das VZS-Vorhaben mündete schließlich in dem Vorhaben „Verteilter Dokumentenserver“ (VDS) der Verbände. Das Ziel des Vorprojektes war es, verkürzt ausgedrückt, das Vorhaben Verteilter Zeitschriftenserver zu initiieren. Dieses ist, formal gesehen, mit der Integration des VDS-Vorhabens in vascoda gelungen. Tatsächlich kann dieses Ziel aber erst dann als vollständig erreicht angesehen werden, wenn die Projekte, welche die AGVerbund für das VDS-Vorhaben vorbereitet (vgl. hierzu den Statusbericht zum VDS/VZS [1] in vascoda), von den Förderern von vascoda bewilligt sind.

Der Schlussbericht konzentriert sich auf den Teil, der den Verteilten Zeitschriftenserver betrifft und der auch das Zentrum der Beratungen und Abstimmungen im Verlauf des hier vorliegenden Projektes bildete. Hier hat das Vorprojekt in Kooperation mit der AG der Verbund-

⁴ für das ZIB/KOBV waren vier Stellen (über 2 Jahre) beantragt, das aber nur als „Vorprojekt“ bewilligt wurde (6 Personenmonate in 2002 sowie Sachmittel für die Evaluation einer Internet-Suchmaschine plus Hardware).

systeme und der GASCO⁵, in der Region Berlin und Brandenburg mit dem FAK⁶, als wesentliches Ergebnis im Berichtszeitraum (bis zum Ablauf des ersten Quartals 2003)

- die Module und Teilprojekte diskutiert, zu denen die am Diskussionsprozess beteiligten Partner erste Projektskizzen⁷ ausgearbeitet haben (Januar 2003).
- eine erste Projektstruktur definiert (März 2003), die in folgende Phasen gegliedert ist
 - I. Spiegel und Speicher, Metadaten und Ressourcen
 - II. Verteilte Zugriffskontrolle und Open Linking
 - III. Nahtlose Navigation und Suche
 - IV. Verteilter Betrieb, Dauerhaftigkeit.
- eine erste technische Evaluation von Suchmaschinen [5] durchgeführt (durch Auftragsvergabe an die Universitätsbibliothek der Universität Bielefeld).

Die Bielefelder Evaluation von Suchmaschinentechnologien führte zur Auswahl der Suchmaschine FAST der Firma Fast Search & Transfer für den Bereich des VZS. Die ausführliche (eigene) technische Evaluation von FAST (durch KOBV/ZIB) in Bezug auf die Realisierung des Verteilten Zeitschriftenservers mit konkretem Content substantiellen Umfangs (Kluwer-Journale) ist zwar in Angriff genommen, jedoch noch nicht vollständig zum Abschluss gekommen. Sie hat aber bereits im Vorprojekt nützliche Ergebnisse für die Überprüfung und Verifikation der konzeptionellen Entwicklung des Verteilten Zeitschriftenservers und des VDS geliefert [3], [4]. Beide Arbeiten sind substantielle Vorarbeiten zum Vorhaben „Virtueller Dokumentenspeicher“ bzw. „VDS des KOBV“ [6], das als das Modell- und Integrationsprojekt des VDS-Vorhabens der AGVerbund anzusehen ist.

Als konkrete Ergebnisse aus der Kooperation mit der GASCO sind ferner für Zwecke der GASCO in Bezug auf das Projekt Verteilter Zeitschriftenserver entstanden:

- Übersichtslisten über Vertragsangebote der Konsortien (Ergebnisse einer Umfrage der GASCO bei den Konsortien).
- Empfehlungen zu Verträgen bzw. zu Ergänzungen von Verträgen hinsichtlich technischer Anforderungen in Bezug auf den Verteilten Zeitschriftenserver.

Diese Übersichten und Empfehlungen enthalten jedoch bestimmte Klauseln aus bestehenden Verträgen und für Verhandlungen mit Verlagen bestimmte Richtlinien, die nicht für die allgemeine Öffentlichkeit gedacht sind, sondern als vertraulich eingestuft wurden. Sie können damit nicht Bestandteil des vorliegenden Berichtes sein.

Zur Rolle der GASCO ist vielleicht generell zu bemerken, dass sie sich als unabhängige (und überregionale) Organisation versteht, die Content-Anbieter und Service-Institutionen – und dazu gehören auch bestimmte vascoda-Partner und die Verbünde – als potentielle Anbieter von Leistungen betrachtet, denen gegenüber der Verhandlungsspielraum voll gewahrt werden muss. Die GASCO will sich, etwas verkürzt formuliert, die Freiheit bewahren, immer den unter Kostengesichtspunkten günstigsten Angeboten am Markt die Präferenz zu geben.

Die Möglichkeit einzelner Konsortien, sich im Einzelfall an die Serviceangebote der Verbundsysteme zu binden oder nicht, bleibt davon unberührt. Bei den Partnern der GASCO sind die Verhältnisse (wie auch bei den Verbundsystemen) recht unterschiedlich. Einige gehören z.

⁵ die „German, Austrian and Swiss Consortia Organisation“.

⁶ dem „Friedrich-Althoff-Konsortium“

⁷ Die Projektskizzen stehen den VZS/VDS-Partnern auf dem Server <http://kirke.hbz-nrw.de/vzs/> zur Verfügung.

B. den Verbänden an, andere nicht. Einige, wie z. B. der Bayerische Bibliotheksverbund oder das überregional operierende Friedrich-Althoff-Konsortium (das nicht mit dem KOBV verwechselt werden darf), unterstützen die Initiative der Verbundsysteme, einen Verteilten Zeitschriftenserver zu errichten, andere haben sich dazu noch nicht in eindeutiger Weise erklärt.

2.3 Konkrete Ziele des Verteilten Zeitschriftenservers

Der Rest dieses Berichtes konzentriert sich auf die Schilderung und kritische Würdigung der technischen Vor- und Grundbedingungen und der technischen Realisierungsmöglichkeiten des Verteilten Zeitschriftenservers.

Die Zielsetzung des Verteilten Zeitschriftenservers lässt sich gemäß dem von der „AG VZS⁸“ der AGVerbund am 26.09.02 beschlossenen Rahmenplan zum VZS wie folgt beschreiben:

„Die Bibliotheken und ihre Partner – Bibliotheksverbände und Konsortien – sehen die Notwendigkeit der Speicherung und des Angebots lizenzierter digitaler Zeitschriften auf eigenen Servern. Die Partner streben mit ihrem Vorhaben in erster Linie die folgenden Ziele an:

- Erhalt und dauerhafte Sicherung einmal erworbener Rechte an digitalen Zeitschriften.
- Bessere Erschließung und Integration digitaler Zeitschriften in das eigene Angebot.
- Nahtlose Navigation in lokalen Zeitschriften-Servern und zwischen digitalen Artikeln.
- Dauerhafte Sicherung des Zugriffs auf digitale Zeitschriften und –perspektivisch– ihre Langzeitverfügbarkeit.

Die Speicherung aller lizenzierten Zeitschriften ist jedoch nur in einer nationalen Kooperation machbar und sinnvoll. Die Partner verfolgen für die Realisierung das vom KOBV vorgeschlagene kooperative Modell. Sie werden schrittweise realisieren:

- Einen Verteilten Dokumentenspeicher mit lokalen Speichern der Verlagsangebote.
- Metasuche, Intranet-Index und Open Linking über diese lizenzierten Zeitschriften.
- Lokale Zugriffs- und Lizenzkontrolle, –perspektivisch– Verteilte Authentifizierung.“

Das besondere Anliegen des ZIB besteht vor dem Hintergrund seiner Math-Net-Aktivitäten und der am ZIB beheimateten KOBV-Zentrale darin, ein Konzept aufzuzeigen, das nicht nur im Bereich der lizenzierten Zeitschriften, sondern für allgemeine – lizenzierte und freie – wissenschaftliche Dokumente Anwendung findet.

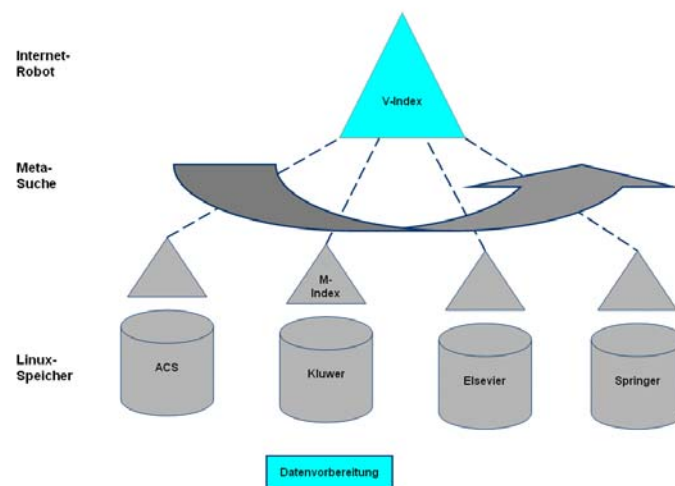
2.4 Verifikation

Die technische Problemanalyse hat das Vorprojekt an Hand einer vereinfachten Modellvorstellung des Verteilten Zeitschriftenservers durchgeführt, in der es noch keine Anbindung an die Metasuche und auch noch kein Open Linking gibt. Diese Modellvorstellung reduziert die verteilte Architektur des VZS auf seinen elementaren Kern, bestehend aus

⁸ die von der AGVerbund eingerichtete Arbeitsgruppe „Verteilter Zeitschriftenserver“; vgl. dazu den Anhang

- einem (im Internet) verteilten Speicher mit oder auch ohne Metadaten (in Abbild 1 durch das Plattenspeichersymbol repräsentiert), der mit möglichst elementaren Technologien realisiert ist (Linux-PC, Internet-Tools),
- einem gemeinsamen zentralen Volltextindex, der durch einen Internet-Robot (Basis: FAST-Suchmaschine) erzeugt wird (V-Index, durch großes Dreieck repräsentiert; in anderen Darstellungen zum VZS wird dafür das umfassende Kreissymbol verwandt),
- einem optionalen, auf Metadaten basierendem Index, der wahlweise dezentral oder auch zentral eingerichtet werden kann (M-Index), falls Metadaten vorliegen.

Die im Rahmen der VZS-Modellvorstellung notwendige Metasuche ist im Schaubild zwar eingezeichnet (geschwungener Pfeil), sie wurde jedoch nicht für Zwecke der Verifikation verwandt. Wir gehen davon aus, dass sich ein lokaler Speicher immer an die Metasuche anbinden lässt, sofern nur Metadaten in einer ausreichenden Struktur vorliegen und der Speicher die Suche in den Metadaten über ein Web-Interface zur Verfügung stellt.



Abbild 2. Vereinfachte Modellvorstellung des VZS

Die Problemanalyse will eine Antwort auf folgende Fragen finden. Ist es prinzipiell möglich und, falls die Antwort positiv ausfällt, unter welchen Bedingungen ist es der Fall,

1. einen kostengünstigen Volltextspeicher digitaler Journale (von einem substantiellen Umfang) mit elementaren technischen Mitteln (Linux-PC, Internet-Tools) so aufzubauen, dass er durch einen Internet-Robot automatisch erschlossen werden kann.
2. den Volltextindex durch einen Internet-Robot (konkret: FAST) automatisch aus Volltextdokumenten zu erzeugen, die von kommerziellen Verlagen geliefert werden.
3. einen Metadaten-Index für den lokalen Speicher mit demselben Internet-Robot (automatisch) zu generieren, sofern vom Verlag auch die Metadaten geliefert werden.

Eine positive Antwort auf diese Fragen würde bedeuten, dass Verbände und Bibliotheken, die sich am VZS mit eigenen Beiträgen beteiligen, auf den Einsatz kostspieliger Spezialsoftware zur Speicherung, Indexierung und zur Verwaltung des „gespiegelten“ Contents einzelner Verlage verzichten könnten. Falls Metadaten zur Verfügung stehen, könnte die Suche in den entsprechenden Beständen mittels der örtlich zur Verfügung stehenden Metasuchmaschine organisiert und auf diese Weise in ein lokales oder regionales Portal integriert werden. Falls diese

aber von einem Verlag (wie z. B. im Falle von ACS) nicht bereitgestellt werden, wäre es – auf dem Wege über die Volltextindexierung – dennoch möglich, den Content im Verteilten Zeitschriftenserver zu erschließen.

Das Vorprojekt zum Verteilten Zeitschriftenserver kann nur eine vorläufige Antwort auf diese Fragen geben. Diese wird im folgenden Teil des Schlussberichts dargestellt und kritisch diskutiert. Eine definitive Antwort ist erst im Zuge der Realisierung des VDS-Vorhabens in vas-coda zu erwarten.

3. Ausgangslage, Grundbedingungen und Kooperative Realisierung

3.1 Ausgangslage

Voraussetzung für die Bereitstellung digitaler Zeitschrifteninhalte (E-Journals) auf Servern der Bibliotheken sind entsprechende Rechte. Elektronische Zeitschriften werden im Regelfall nicht gekauft, sondern für einen vertraglich vereinbarten Zeitraum auf einem Server nach Wahl des Anbieters (Verlage) für Nutzung und Zugriff lizenziert. Das Recht, einmal bezahlte Zugangslizenzen für Zeitschriftenjahrgänge auch nach Ende eines laufenden Lizenzvertrages oder nach Ende der Bereitstellung durch den Anbieter zu nutzen, wird von den vertragführenden Bibliotheken, Konsortien und Informationsverbänden – seit 2001 verstärkt – als wesentlicher Bestandteil der Lizenzverträge in die Verhandlungen mit den Anbietern einbezogen.

Das Recht zur Speicherung dieser Daten auf einem Server ihrer Wahl und unter Beibehaltung des vereinbarten Lizenzrahmens wird ebenfalls von den vertragführenden Bibliotheken und Konsortien angestrebt. Das Hosten von digitalen Zeitschriften auf leistungsfähigen Servern mitten im Wissenschaftsnetz liegt im Interesse der Verlage selbst. Dieses spiegelt sich auch in der rechtlichen Situation im Bereich des Friedrich-Althoff-Konsortiums wider, die sich einer Darstellung (vom Dezember 2003) von M. Rieck, dem gegenwärtigen Vorsitzenden des FAK, folgend kurz gefasst so darstellt:

- ACS: Das Hosten der Volltexte ist vertraglich abgesichert. Die Lieferung der PDF-Dateien erfolgt inzwischen (laufend) an die UB der Humboldt-Universität zu Berlin „bare bone“ (ohne Metadaten und Navigation). Eine Vereinbarung im Bayerischen Verbund mit ACS beinhaltet das Hosten der PDFs durch den KOBV, also über Verbund- und Ländergrenzen hinweg.
- Elsevier: Die Möglichkeit des Vorhaltens von Elsevier-Daten ist (laut einem Schreiben des Verlages an das FAK) Bestandteil der Vertriebspolitik.
- Kluwer: Das Vorhalten der Volltexte ist vertraglich abgesichert. Die Lieferung erfolgt (einschl. der Metadaten) inzwischen laufend an das ZIB.
- Springer: Die Verfügbarkeit der Archivdaten ist vertraglich abgesichert, jedoch derzeit noch nicht realisiert (weil eine Frage der Kosten).

Das Recht zum Anbieten dieser Inhalte auch an Dritte, d.h. über den Kreis der Vertragspartner hinaus, wurde in Parallele zu Fernleih- und Dokumentlieferverfahren bisher als selbstverständlich vorausgesetzt, was es aus Sicht der Verlage aber nicht (immer) ist. Die Verträge zu den projektrelevanten Zeitschrifteninhalten, die für die Antragsteller über die Einkaufsgemeinschaft z.B. des Friedrich-Althoff-Konsortiums (FAK) abgeschlossen wurden, enthalten im Hinblick auf das Projektthema folgende Optionen:

- bei den Anbietern ACS, Blackwell Publishing, Elsevier/Academic Press, Kluwer Academic Publishers und Springer ein Recht zur Nutzung der Daten der Vertragsjahre nach Vertragsende,
- bei den Anbietern ACS, Elsevier/Academic Press, Kluwer Academic Publishers und Springer ein Recht auf Auslieferung der Daten zur Übernahme auf einen Server eigener Wahl.

Ein dringender Bedarf an Zugriffsoptionen auf diese Inhalte (Zeitschriftenjahrgänge), weil Verträge entweder beendet wurden oder sich das Titel- und Teilnehmerspektrum gewandelt hat, besteht für jeweils mehrere Erscheinungsjahre für die Anbieter Elsevier/Academic Press und Springer und für den Anbieter ACS.

3.2 Besondere Erfahrungen des KOBV und des ZIB, Vorbedingungen

Der Kooperative Bibliotheksverbund Berlin-Brandenburg (KOBV) ist eine Einrichtung der Länder Berlin und Brandenburg und ihrer Bibliotheken. Er wurde in den Jahren 1997 bis 2001 am Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB) in einem von den Ländern Berlin und Brandenburg geförderten Projekt aufgebaut und am 20. August 2001 institutionalisiert. Die KOBV-Zentrale hat ihren Sitz im Land Berlin und ist in das ZIB eingegliedert. Sie ist im Bereich Scientific Computing, Abteilung „Wissenschaftliche Informationssysteme“ angesiedelt. Leiter des Bereiches ist Prof. Dr. Martin Grötschel. Die Leitung der KOBV-Zentrale wird durch den Leiter der Abteilung Wissenschaftliche Informationssysteme wahrgenommen. Kooperationspartner im KOBV sind Bibliotheken aus der Region Berlin-Brandenburg.

Das KOBV-Projekt hat in der Projektphase mit der Entwicklung der KOBV-Suchmaschine die technischen Grundlagen für die heutige Institution KOBV gelegt. Es setzte dabei konsequent auf die neuen Technologien des Internet und des World Wide Web, insbesondere auch auf international standardisierte Protokolle (Z39.50, http). Die KOBV-Suchmaschine realisiert mit Mitteln der Metasuche (parallel verteilte Suche auf der Basis von Z39.50 und http) einen virtuellen Verbundkatalog und realisiert gleichzeitig (und über die Metasuche ansprechbar) einen MetaIndex (durch Zusammenspielen gemeinsamer Katalogdaten). Beide Arten, einen virtuellen Verbundkatalog zu bilden, haben sich als sinnvoll und auch als notwendig erwiesen. Mit beiden stellt der KOBV seinen Nutzern ein modernes Informationswerkzeug zur Verfügung, mit dem sie unabhängig von Zeit und Ort im Internet in den wesentlichen Bibliotheksbeständen der Region suchen können: in allen Universitäts- und Fachhochschulbibliotheken, in der Staatsbibliothek zu Berlin, in vielen Spezialbibliotheken, in den öffentlichen Bibliotheken Berlins und Brandenburgs und – über die Region hinaus – in allen deutschen Bibliotheksverbänden.

Die technische Konzeption zur Realisierung des vom KOBV für den Verteilten Zeitschriftenserver vorgeschlagenen Modells ist im Wesentlichen eine Verallgemeinerung der Idee der KOBV-Suchmaschine. Der Grundgedanke besteht darin, ähnlich der Realisierung eines Virtuellen Katalogs (aus bestehenden „realen“ Katalogen) mit Mitteln der Metasuche und des Metaindex aus bestehenden „realen“ Speichern einen Virtuellen Speicher zu realisieren.

Im Einsatz der Metasuche – konkret, der gemeinsam mit der Firma Ex Libris Ltd. entwickelten Metasuchmaschine MetaLib – verfügt die KOBV-Zentrale über hinreichende Erfahrungen. Die Vorteile der Metasuche lassen sich kurz wie folgt skizzieren:

- Die Integration neuer Systemtypen (von Speichersystemen) ist möglich – dafür sind geeignete Metadaten notwendig, die über ein Web-Interface abfragbar sind.
- Ist ein bestimmter Systemtyp einmal angeschlossen, lassen sich weitere Systeme desselben Typs relativ einfach in die Metasuche einbeziehen.
- Ein auf diese Weise integriertes System ist technisch weniger effizient (als ein gemeinsamer Index) und in einem gewissen Sinne auch fragiler (erhebliche Leistungsanforderung an das kleinste System).
- Nutzer erhalten die Möglichkeit zur Auswahl von Zielsystemen.
- Es ergeben sich neue Möglichkeiten zur Personalisierung und Zugriffsteuerung.

Im Vergleich zum Metaindex, der auf einem integrierten Bibliothekssystem basiert und in Bezug auf die Integration von Daten aus von diesem unterschiedlicher Bibliothekssysteme recht schwierig und nur sehr aufwendig zu handhaben ist, kann die Metasuche außerordentlich flexibel und leicht erweiterbar eingestuft werden: Die KOBV-Suchmaschine ist z.B. heute in der Lage, die OPACs praktisch aller im Bereich des deutschen Bibliothekswesens vertretenen Bibliothekssysteme in ihre Metasuche zu integrieren.

Dennoch skaliert die Metasuche nicht (im Sinne des Internet und des Web). Die Anbindung eines jeden einzelnen Systems ist mit einem erheblichen Umfang an „Handarbeit“ verbunden und der Betrieb eines virtuellen Kataloges von nur mäßigem Umfang (unter 100 Systemen) bedarf der ständigen und intensiven Beobachtung und Überwachung. Sie allein – dies ist vielleicht die wichtigste Erfahrung des KOBV-Projektes – reicht nicht aus. Ein effizienter, stabiler gemeinsamer Index ist ebenso notwendig. Dieser müsste allerdings mit wesentlich weniger Aufwand zu erstellen sein, als das beim Gemeinsamen KOBV-Index notwendig war.

Die Arbeitshypothese des vom KOBV vorgeschlagenen Modells ist, dass sich – bei Vorliegen von Volltexten – der gemeinsame Index des Verteilten Zeitschriftenservers mit den Mitteln eines Internet-Robots automatisch bilden lässt. Auf diese Weise werden sich Dokumente in einem verteilten Speicher auf „virtuellem Wege“ gemeinsam organisieren (und erschließen) lassen. Die These ist, dass ein Internet-Index als Mittel der Integration eines verteilten Dokumentenspeichers mit folgenden Eigenschaften einhergeht:

- Niedriger Erstaufwand für die Integration neuer Systemtypen (von Speichersystemen) – die einzelnen Dokumente müssen dabei über Web-Navigation zugänglich sein.
- Weitere Zielsysteme desselben Typs lassen sich mit automatischen Mitteln (Robotbasiert) integrieren. Die Methode skaliert.
- Ein auf diese Weise gebildeter Index arbeitet effizient und stabil.
- Die Suche erfolgt (allerdings) immer im Gesamtbestand, der mit dem Internet-Robot adressiert wurde. Dieser lässt sich auf – ausschließlich – relevante Bestände steuern.
- Ein Internet-Index lässt sich in lokale Portale (Metasuchmaschinen) integrieren.

Inzwischen liegen durch mehrere im World Wide Web operierende Internet-Suchmaschinen hinreichende Belege für die Plausibilität dieser Arbeitshypothese vor. Die Auswahl der konkreten FAST-Suchmaschine für den VZS wurde im Rahmen einer Evaluation von Suchmaschinen durch die UB Bielefeld getroffen [5]. Das ZIB verfügt ferner im Math-Net-Projekt über hinreichende Erfahrungen mit dem Sammeln („Harvesting“) von Volltexten.

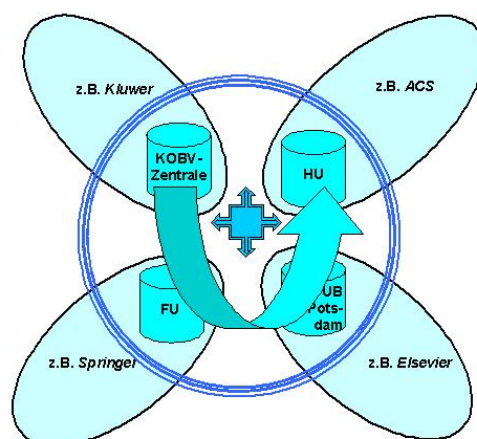
3.3 Zielsetzung des Verteilten Dokumentenspeichers des KOBV

Zielsetzung des VDS des KOBV sind Speicherung, Erschließung und dauerhafte Verfügbarkeit digitaler Inhaltsressourcen der am Projekt beteiligten Bibliotheken. Mit Priorität wird angestrebt, dem Nutzer eine integrierte Suche und nahtlose Navigation über den Bestand der elektronischer Zeitschriften in der Region anzubieten. Hierzu gehören die folgenden Dienste.

- Die Metasuche über alle elektronischen Zeitschriften.
- Die Volltextsuche über alle elektronische Zeitschriften.
- Der Online-Zugriff auf Dokumente mit Lizenzen der jeweiligen Bibliothek.
- Das Angebot weiterer Optionen: Dokumentenlieferung, Fernleihe, Pay-Per-View etc.
- Die Navigation über Gesamtbestand und lokale Speicher der Projektbibliotheken.
- Die nahtlose Navigation mittels gültiger Referenz-Links in Dokumenten.

Der verteilte Dokumentenspeicher soll auf der Basis erworbener Zugangslizenzen aktuelle und zurückliegende Zeitschriftenjahrgänge sowie frei zugängliche Bestände und Sammlungen (Dissertationen, Tagungsberichte, Monographien, E-Learning-Contents, Bilder etc.) unter einer einheitlichen integrierenden Oberfläche umfassen. Es wird angestrebt, den Zugang zu lizenzierten Produkten nicht nur über IP-Kontrolle, sondern über eine verteilte Authentifizierung zu ermöglichen (Zugriff von „zu Hause“ aus). Der verteilte Dokumentenspeicher bringt das digitale Dokumentenangebot der Bibliotheken direkt an den Arbeitsplatz. Die Bibliotheken (des KOBV) werden ihren Auftrag der gesicherten Bereitstellung elektronischer Informationsressourcen auf eine zeitgemäße Art und Weise gewährleisten können.

Als Modellvorstellung wird die Realisierung eines verteilten Dokumentenspeichers in einer Form favorisiert, die zu Beginn des Jahres 2002 für die Speicherung digitaler Dokumente im KOBV entworfen wurde [7], [8]. Die Grundidee dieses Konzeptvorschlages ist es, die digitalen Zeitschriften der Verlage oder Teile davon in lokalen Spiegeln der Verlage oder in Dokumentenspeichern mit verlagsunabhängiger Software zu kopieren (Abbild 3).



Abbild 3. Modell des VDS des KOBV

Eine Metasuche (geschwungener Pfeil) verbindet hier die lokalen Spiegel und Dokumentenspeicher. Eine Internet-Suchmaschine (Robot-basiert) erschließt automatisch den gesamten Bestand durch einen effizienten gemeinsamen Volltextindex bzw. Intranet-Index (Kreissymbol). Ein Linking-Service (Pfeilsymbol in der Mitte) regelt auf der Grundlage von OpenURLs

den Zugriff auf Dokumente, wobei er Nutzer ohne Zugriffsberechtigung zu Lieferdiensten, wie Fernleihe, Pay-per-view etc., umlenkt.

Die Modularität des Konzeptes macht eine Verteilung der Aufgaben auf viele Schultern möglich. Denkbar wäre nach Vorstellung des KOBV– auf nationaler Ebene – etwa die folgende Aufgabenteilung :

Intranet-Index / Robot	⇒ Kompetenzzentrum (regional, ggf. auch national)
Metasuchmaschine	⇒ Verbund (regional)
Lokaler Speicher / Spiegel	⇒ Verbund (regional) / Bibliothek (lokal)
Linking Server	⇒ Verbund (regional) / Bibliothek (lokal)
Zugriffskontrolle	⇒ Verbund (regional) / Bibliothek (lokal)
Verlagsserver	⇒ Verlag / Host (international)
Verträge / Erwerbung	⇒ Konsortium / Bibliothek

Das technische Konzept „skaliert“ im Sinne der Kooperation. Neue Bibliotheken können leicht hinzutreten, auch mit eigenen digitalen Angeboten oder E-Journalen aus dem IuK-Bereich. Für die Startphase wäre aber im Interesse der Machbarkeit (und schlanken Organisation) zu überlegen, die Hauptlast den Verbundzentren und großen (leistungsfähigen) Bibliotheken zu übertragen.

3.4 Grundbedingungen, Kooperative Modellvorstellung

Die Grundprinzipien der vom KOBV den Verbänden vorgeschlagenen kooperativen Vorgehensweise sind die Prinzipien (und Standards) des Internet und des World Wide Web:

- Offenheit
- Heterogenität
- Dezentralität

Sie sind in den Publikationen des KOBV hinreichend ausgeführt und kritisch diskutiert (siehe die auf den Web-Seiten des KOBV, www.kobv.de, zur Verfügung stehenden Publikationen zum Projekt „KOBV-Suchmaschine“). An dieser Stelle sollen zwei weitere grundsätzliche Annahmen und Eigenschaften diskutiert werden, die in die Architektur der generischen Modellvorstellung des KOBV-Dokumentenspeichers eingegangen sind und die auch in den technischen Diskussionen der Arbeitsgruppen des Verteilten Zeitschriftenservers immer wieder reflektiert wurden:

- Skalierbarkeit
- Nahtlose Navigation

Skalierbarkeit, manchmal auch (uneingeschränkte) Erweiterbarkeit genannt, ist eine der zentralen Forderungen, die an die technischen Konzeptionen – Protokolle, Implementierungen und Standards – des Internet und des World Wide Web gestellt werden. Ist ein technisches Konzept nicht skalierbar, ist es also nicht auf eine sehr große Zahl von Teilnehmern, Servern, Speichern etc. erweiterbar, so scheidet es für den Gebrauch in einem weltweiten Netz von vornherein aus. Eine der wichtigsten technischen Folgerungen, die sich aus der Skalierbarkeit ergeben, ist die automatische Erzeugung von zentralen Netzwerk-Eigenschaften. Skalierbarkeit wird deshalb auch immer wieder mit Begriffen wie Offenheit und Interoperabilität in Verbindung gebracht. Ein Netz, das nur sehr mühsam und mit viel Handarbeit zu knüpfen ist, wird in der Netz-Community als nicht offen abgelehnt.

Die Einstiegsschwelle in ein Netzwerk muss möglichst niedrig sein. Eine möglichst weitgehende Automatisierbarkeit wird angestrebt.

Das beste Beispiel für Skalierbarkeit eines Konzeptes ist vielleicht der Link bzw. die Adresse im Netz. Es genügt ein einziger Link (aus dem World Wide Web) zu einer Homepage, um sie – automatisch – zu einem Bestandteil des World Wide Web zu machen. Suchmaschinen im Web können nun die Homepage über diesen Link automatisch finden und integrieren. Ein anderes (vor den Zeiten des Web bestehendes Konzept) ist der Server. Es genügt z.B., die Adresse eines ftp-Servers zu kennen, um auf seinen gesamten Bestand mit automatischen Mitteln zugreifen zu können. Auf der Anbieterseite reicht ein einfacher Kopierprozess in das hierarchische Filesystem eines ftp-Servers aus, um etwa ein Dokument einer weltweiten Gemeinde (im Sinne einer großen Anzahl) an Nutzern des ftp-Dienstes zum automatischen Abruf zur Verfügung zu stellen. Das ftp-Konzept (Protokoll, Server und Client) skaliert. Sowohl die Anbieter- als auch die Abnehmerseite sind einfach zu bedienen. Skalierbare Konzepte ermöglichen die Kommunikation zwischen einer sehr großen Zahl von Teilnehmern (Anbietern/Nutzern).

Die Nutzung international verbreiteter Standards ist genau deshalb die zentrale Voraussetzung und Forderung von und in Netzwerken. Komponenten, die auf proprietären Standards aufbauen oder sie voraussetzen, scheiden für den uneingeschränkten Gebrauch in offenen Netzen aus. Sie skalieren nicht.

Leider gibt es dieses „Fehlverhalten“ (im Sinne des Aufbaus von Netzwerken) nicht nur im Bereich der Programme und Protokolle, sondern auch bei den Datenstrukturen. Unterschiede in den Datenstrukturen führen i. d. R. dazu, dass Kommunikationsbeziehungen nur schwer zu etablieren sind, Unterschiede in den Dateninhalten ebenfalls. Ziel beispielsweise der Dublin-Core-Initiative ist es, diese Unterschiede zu überwinden und Interoperabilität zu erreichen. Aus verschiedenen Gründen⁹ ist der Datenüberbau¹⁰ der verschiedenen Publikationsserver zu unterschiedlich, als dass man die zugrunde liegenden Dokumente einfach in einem großen Pool zusammenführen und automatisch integrieren könnte. Das wird nicht einmal dann möglich sein¹¹, wenn man nur die in Frage kommenden Metadaten und Navigationsstrukturen betrachtet. Dieses ist der Grund dafür, dass der Verteilte Dokumentenspeicher des KOBV und der verteilte Zeitschriftenserver von verlagsspezifischen Spiegeln oder Speicherservern ausgehen.

Unter dieser Perspektive wird ein Problem offensichtlich, das in der WWW-Community unter dem Stichwort Uniform Resource Name zwar bekannt, aber nicht allgemein gelöst wurde, insbesondere nicht für den Fall der „richtigen“ Selektion von Dokumenten¹². Im Bereich der wissenschaftlichen Publikationen, sowohl der lizenzierten als auch der freien Materialien, tritt dieses Problem immer dann auf, wenn es alternative Quellen oder Server gibt. Das ist z. B. immer dann der Fall, wenn eine Bibliothek die Lizenz bei einem Aggregator in Anspruch nimmt oder wenn es, wie es bei freien digitalen Journalen üblich ist, über die ganze Welt verstreute Spiegelserver gibt. Größere Verlage führen ebenfalls selbst mehrere Spiegel ihrer Verlagsprodukte.

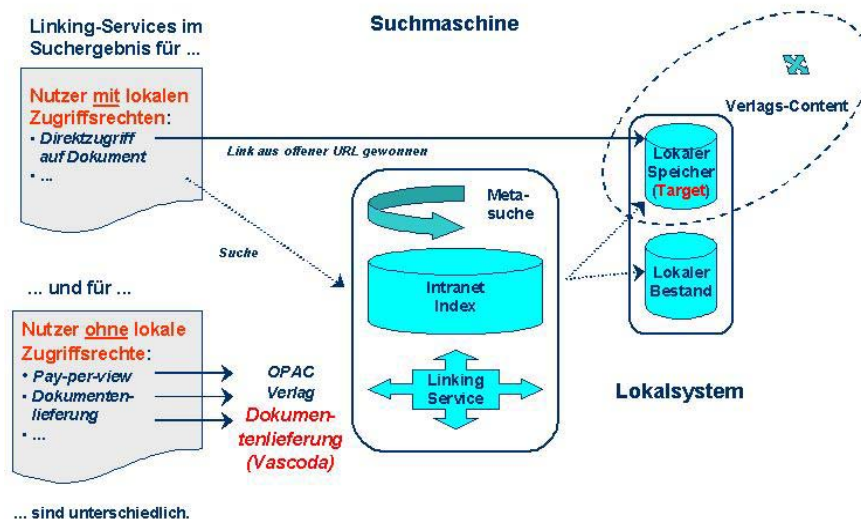
⁹ Mangelnde Erfahrung oder auch ein spezifisches Interesse, sich am Markt in besonderer Weise zu profilieren.

¹⁰ Metadaten, Navigationsstrukturen, Organisations- und Statistikinformation. Nutzerdaten und Sicherheit.

¹¹ Dies ist jedenfalls die Einschätzung der Arbeitsgruppe „Dokumentenspeicher im KOBV“

¹² Im Bereich der digitalen Journale tritt dieses Problem immer dann auf, wenn es neben der ersten Kopie eines Dokumentes im Verlagsserver weitere Kopien gibt.

Auf welchen Spiegel oder Speicher wird eine Suchmaschine oder eine Bibliothek nun in ihrem Portal verweisen? Sie wird ihre Nutzer per URL dorthin lenken wollen, wo sie über Lizenzen verfügt¹³ und im anderen Falle zur Fernleihe oder zur Dokumentenlieferung.



Abbild 4. Kontextsensitive Steuerung des Zugriffs auf Dokumente

Der Vorschlag des KOBV ist, dieses Problem mittels kontextsensitivem Linking bzw. Open-URL-Management zu lösen, wobei die Steuerung des Benutzers durch die lokale Bibliothek selbst erfolgt. Nur sie hat eine vollständige Übersicht über ihre Lizenzen, nur sie weiß über die Rechte ihrer Nutzer Bescheid und nur sie hat den Zugriff auf die Daten ihrer Nutzer.

Mittels kontextsensitivem Linking in ihrem lokalen Portalsystem kann die lokale Bibliothek ihren Nutzern auch noch dann – weitere – Hilfestellung geben, wenn sie nicht über die Lizenz auf ein von ihm gefundenes Dokument verfügt. Sie kann ihn etwa zu einem (für ihn) günstigen Dokumentenlieferdienst lenken oder, was bei einem Naturwissenschaftler heute der Regelfall ist, ihn durch das Angebot eines entsprechend konfektionierten Links zu einer Suche nach dem Dokument (oder auch nur dessen Autor) im Web veranlassen.

4. Die vier Phasen des Verteilten Dokumentenservers

Dieses Kapitel diskutiert die Module und Komponenten des VDS¹⁴ aus technisch-organisatorischer Sicht in Form von vier Phasen. Der Schlussbericht ergänzt die einzelnen Phasen jeweils um Abschnitte „Verifikation“, in denen die Zielvorstellungen der jeweiligen Phase konstruktiv und kritisch hinterfragt werden.

Phase 1 - Spiegel und Speicher

Der Aufbau und Betrieb von Spiegeln und Speichern ist die Sache der Verbünde und großen leistungsfähigen Bibliotheken, die das Angebot von Content und die entsprechende Unterstützung der ihnen zugehörigen Bibliotheken zu ihrem Zuständigkeitsbereich zählen. Die Einbin-

¹³ Sonst müsste der Nutzer „sein“ Dokument noch einmal bezahlen.

¹⁴ Im Folgenden diskutieren wir nur noch den VDS, der allerdings den VZS umfasst.

dung der Speicher und Spiegel in ihre Metasuchmaschinen und Portalsysteme wird ebenfalls von den Verbundsystemen, die Portalsoftware einsetzen (und z. T. selbst mit entwickelt haben), als eine ihrer Betriebsaufgaben gesehen.

Die Ausarbeitung von entsprechenden Verträgen oder Vertragsmodellen ist Sache der Konsortien, die in einigen Fällen mit den Verbänden „deckungsgleich“ sind. Einige Verbände und Verbundbibliotheken betreiben bereits Spiegel von Verlagsangeboten im Routinebetrieb, etwa

- HeBIS Elsevier/Academic-Spiegel (Science-Direct-Software)
- HBZ Elsevier-Spiegel (Science-Direct-Software)
- UB-Bielefeld Academic-Spiegel (eigene Retrieval-Software)

Andere Verbundsysteme planen den Aufbau. In Vorbereitung sind

- KOBV ACS-Speicher (eigener Dokumentenspeicher)
Elsevier-Speicher (basierend auf Internet-Tools und Linux)
Kluwer-Speicher (basierend auf Internet-Tools und Linux)
Springer-Speicher (DSpace-Software des MIT)
- UB-Freiburg Lippincott-Spiegel (Ovid-Software)

Die BSB hat angekündigt, sich am VDS ebenfalls mit einem Speicherserver zu beteiligen.

Nicht alle Verbundsysteme und Bibliotheken verfügen jedoch über entsprechende Erfahrungen oder Content-Management-Systeme. Sie bedürfen beim Aufbau entsprechender Infrastruktur der Unterstützung durch Förderorganisationen. Bei der Vorbereitung entsprechender Betriebsaufgaben ist ferner auch der „Langzeitaspekt“ zu berücksichtigen. Die Langzeitar Archivierung ist heute jedoch noch als F&E-Projekt anzusehen.

Der KOBV plant, sein Content-Angebot unter dem Aspekt der Dauerhaftigkeit der Bereitstellung als einen verteilten Content-Speicher aufzubauen. Er hat zur Unterstützung der Realisierung eines Virtuellen Dokumentenspeichers (VDS des KOBV) einen DFG-Antrag gestellt [6].

Verifikation

Die Spiegelung und Speicherung – das Hosten – von Verlags-Content ist eine relativ neue Aufgabe für Bibliotheken und Verbände. In einigen Fällen kann hierfür verlagseigene Software eingesetzt werden, die dann gut auf den Content abgestimmt ist. Abgesehen vom in Frage kommenden Umfang der Daten, der zugehörigen Metadaten und Navigationsstrukturen treten hierbei keine prinzipiellen Probleme auf. Viele Verlage können jedoch keine „eigene“ Software zur Verfügung stellen. Sofern sie keinen eigenen Server betreiben, vertreiben sie ihren Content dann i. d. R. über einen Aggregator, wie z. B. über OVID. Content-Systeme der Aggregatoren sind jedoch hochspezialisiert und i. A. nicht käuflich zu erwerben.

Erfahrungen des ZIB/KOBV mit der Speicherung der Kluwer-Daten des FAK [3] zeigen, dass für das Hosten großer Datenbestände nicht unbedingt ein kommerzielles Content-System zum Einsatz kommen muss. Bei geschicktem Einsatz von Linux-basierter Software, Apache-Server und Internet-Tools sind vergleichbare Funktionen relativ kostengünstig zu realisieren, sofern für Navigation und Suche ein Portalsystem, das „vorgeschaltet“ wird, zur Verfügung steht.

Die erkennbaren Probleme beim Aufbau des Kluwer-Spiegels lagen eher im organisatorischen als im technischen Bereich. Sie betreffen Prozeduren, die im Binnenverhältnis mit dem Verlag vertraglich abgestimmt und festgelegt werden müssen, z. B. bezüglich der

- Vollständigkeit und Konsistenz der Datenlieferung
- Integrität des Speicherservers (haben sich Fehler eingeschlichen?)
- Konformität der Daten mit den PDF-Spezifikationen
- Erkennung und Korrektur von Fehlern im Datenbestand

Im Gegensatz zu einigen Verlagen ist Kluwer in der Lage, zu den Volltexten auch Metadaten in strukturierter Form zu liefern. Es wäre eine technische Herausforderung, ein automatisches Verfahren zu entwickeln, das aus PDF-Files (also aus Druckformaten) unterschiedlicher Variationsbreiten in Bezug auf den PDF-Standard Metadaten extrahieren könnte.

Die Verfügbarkeit von Metadaten könnten ein weiteres Problem darstellen. Bei Kluwer ist dieses jedoch nicht der Fall, denn sie sind Bestandteil der Datenlieferung. Generell gilt: Falls Metadaten vorhanden sind, sollten diese mit automatischen Mitteln zu einem Index zusammengeführt werden können, der dann über ein Web-Interface zur Verfügung steht. Dass diese Aufgabe mit einem Internet-Robot erledigt werden kann, wird in der Darstellung zur Phase 3 geprüft und dargelegt. Falls ein Verlag keine Metadaten liefern kann oder will oder diese aus anderen Gründen nicht geeignet sind, kann ein Robot aus den Volltexten einen Volltextindex aufbauen, der dann „ersatzweise“ auf dem lokalen Speicher zur Verfügung steht. Fragen dieser Art werden in dem Erfahrungsbericht zum Einsatz des FAST-Systems behandelt [4].

Phase 2 – Verteilte Zugriffskontrolle, Open Linking

Die Grundbedingung für die Akzeptanz einer solchen Architektur, die auf Metasuche, Intranet-Index und Open Linking für lizenzierte Materialien beruht, besteht bei Bibliotheken, Verlagen und Nutzern darin, dass (zumindest) ein „Schaufenster“ (Ergebnislisten der Suchoperationen der Nutzer) auf die von den Bibliotheken angebotenen Inhalte in den Portalen der Bibliotheken und des VDS sichtbar ist und der Zugriff auf die Volltexte denjenigen Nutzern ermöglicht wird, deren Bibliotheken über die entsprechenden Lizenzen verfügen. Auf diese Weise können nicht nur die berechtigten Interessen der Verlage gewahrt werden (Vertrauensschutz), sie erhalten gewissermaßen im Gegenzug mit einem Spiegel- oder Speicher-Server selbst ein „leistungsfähiges Schaufenster“ im Wissenschaftsnetz, dessen dauerhafte Funktionsfähigkeit von den beteiligten Bibliotheken getragen wird.

Im Falle eines Spiegelservers bleiben sogar die originären und oft sehr leistungsfähigen Mechanismen der Navigation und Suche des Content-Anbieters (wie etwa im Falle des Einsatzes von Ovid-Software) erhalten. Nutzer, deren Bibliothek über die entsprechende Nutzungslizenz verfügt, können aus den Ergebnislisten der Suche direkt zum gefundenen Zieldokument springen, das Dokument dort abrufen oder auch von dort mit den Mitteln des Zielservers weiter navigieren. Der Server des Verlages wird durch den VDS nicht etwa abgeschafft. Der Verteilte Zeitschriftenserver steht mit seinen Such- und Navigationsinstrumenten nicht „über“ den Servern der Verlage und Agenturen, sondern eher „zwischen“ ihnen. Mit seinen relativ einfachen, aber sehr allgemeinen Mechanismen will er dem Nutzer einen einfachen, universellen Zugang schaffen und dabei zwischen den verschiedenen Content-Angeboten „nur“ vermitteln.

Die Bedingungen einer solchen Navigation und Suche in den Beständen müssen natürlich in die bestehenden Nutzungsverträge eingearbeitet werden und dort auf Dauer verankert sein. Werden Verlage dazu bereit und willens sein? Es besteht bei den bei Konsortien und großen Bibliotheken bereits bestehenden Verträgen Grund zur Annahme, dass sich diese in der skizzierten Form allgemeiner fassen und vereinbaren lassen. Manche Verlage gehen sogar noch weiter und stellen Bibliotheken ihre Metadaten (nicht nur lokale Suchergebnislisten) zur Verfügung. Einige Verlage sind auch dazu bereit, ihre Indexdaten für dieselbe Form der Nutzung bereitzustellen. In allen diesen Fällen muss allerdings gesichert sein, dass nur die berechtigten Nutzer Zugriff auf die Dokumente bzw. Volltexte erhalten.

Die Schlussfolgerung ist, dass für den Aufbau eines VDS die Entwicklung einer verteilten Zugriffskontrolle, deren Funktionsfähigkeit von den beteiligten Verlagen zertifiziert wird, eine Schlüsselfunktion ist. Der Aufbau und die Wahrung einer Vertrauensbeziehung zu den Verlagen ist die notwendige Voraussetzung für den Aufbau und Betrieb des VDS. Die technische Architektur der Zugriffskontrolle kann auch nur dezentral, heterogen und damit verteilt angelegt sein, da die entsprechenden Daten über Zugriffsrechte und Rollen der Benutzer (die Fragen der Authentifizierung und Autorisierung betreffen) nur bei den lokalen Bibliotheken vorliegen.

Was geschieht in einem solchen System mit einem Benutzer, dessen Bibliothek über keine Lizenz zur Nutzung und zum Zugriff auf ein gefundenes Dokument verfügt? Eine allgemeine Antwort wäre, dass solche Nutzer zum Server des Verlages oder zu einer anderen Art der Dokumentenlieferung „umgelenkt“ werden sollten. Heute, in den Zeiten des Internet, sind aber ganz andere Präferenzen sinnvoll und möglich. Das Blue Ribbon Advisory Panel on Cyberinfrastructure der NSF eröffnet in ihrem Planungspapier „Revolutionizing Science and Engineering Through Cyberinfrastructure“ vom Januar 2003 die folgende Perspektive:

“The primary access to the latest findings in a growing number of fields is through the Web, then through classic preprints and conferences, and lastly through refereed archival papers.”

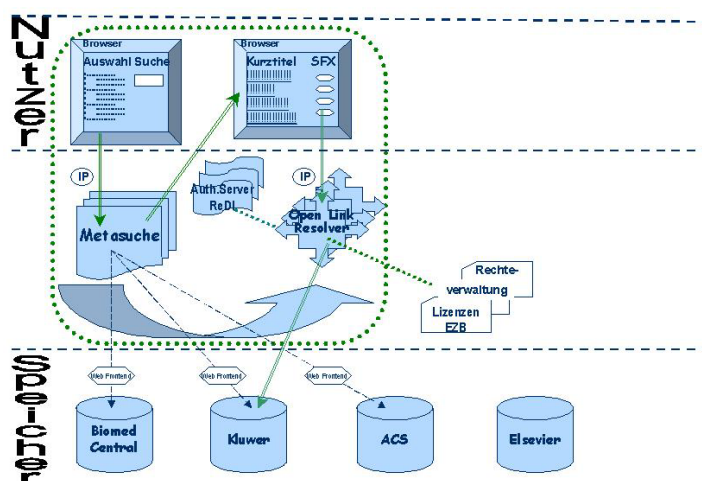
Die „Beste Vorgehensweise“ gibt es nicht mehr. Was zu tun ist, hängt von mehreren Faktoren ab: vom Fach, von der beruflichen Position des Benutzers, von den IT-Kenntnissen des Nutzers, von der Geschwindigkeit, mit der seine Bedürfnisse zu erfüllen sind, von den Mitteln, über die er persönlich oder über die seine Bibliothek verfügt und natürlich auch von den Lizenzen seiner Bibliothek, seines Fachbereiches oder seiner Arbeitsgruppe. Einige Beispiele sollen die heute gegebene Variationsbreite verdeutlichen:

- In einigen Fachwissenschaften, wie z.B. der Mathematik, Physik oder Informatik kann man zu einem gefundenen Titel dieselbe Publikation (oder neuere) im Netz finden.
- Die Suche nach der Homepage des Autors im Web ist sicherlich eine gute Stelle, verwandte oder aktuellere Arbeiten, Preprints oder Publikationen von Workshops zu suchen. Diese findet der Benutzer, wenn ihn die Bibliothek mit den im VDS gefundenen (Meta-) Daten zu einer Google-Suche umlenkt.
- Nicht alle Fachbereiche oder Bibliotheken sind in der Lage, allen ihren Mitgliedern und Studenten den vollen Zugriff auf Fachdatenbanken zu bieten. Nicht alle Studenten verfügen über hinreichende Kenntnisse in der Bedienung von Fachdatenbanken.
- Wissenschaftliche Publikationen hängen oft mit Primärdaten, Visualisierungen, Foto- und Videodokumenten etc. zusammen, zu denen Bibliotheken ihre Nutzer – etwa ausgehend von den Metadaten einer erfolgreichen Suche im VDS – durch das Angebot entsprechender Suchoptionen umlenken können.

Die technische Methode, von einem Dokument zu relevanten Kontexten zu kommen, ist schon seit längerem in der Hypertextforschung unter dem Stichwort „Context Sensitive Linking“ bekannt, hat jedoch noch nicht ihren Platz im World Wide Web gefunden (vgl. dazu auch den Hauptvortrag „The Missing Link“ von Wendy Hall zur WWW2003 Jahreskonferenz in Budapest). Im Bereich der Bibliotheken ist allerdings eine spezielle Form des Context Sensitive Linking unter den Stichworten „OpenURL“ und „Open Linking“¹⁵ seit einiger Zeit bekannt. Es ist die Methode der Wahl für Bibliotheken, ihren Nutzern in Abhängigkeit von ihren speziellen Bedingungen alternative Suchoptionen auf der Basis von vorhandenen bzw. in Form von OpenURL kodierten Metadaten anzubieten.

Open Linking ist auch eine Methode für Bibliotheken, URLs in und zu den Dokumenten im VDS korrekt einzustellen, d. h. so zu konfigurieren, dass sie den Benutzer nicht mehr (automatisch) zum Server des Verlages führen, sondern zum entsprechenden Spiegel- oder Speicherserver im VDS. Die Voraussetzung für diese Art des Managements von Links ist, dass die Links im Content als OpenURL ausgeführt sind. OpenURLs und mit diesen „kooperierende“ DOIs¹⁶ werden an Stelle von „einfachen“ URLs von den Datenbank- und Content-Anbietern eingesetzt, um von Referaten, Abstracts und Literaturziten korrekte Links (gelegentlich auch „Tiefenlinks“ genannt) in die Dokumente zu schaffen. Für Server, die nicht OpenURL-enabled sind¹⁷, können OpenURLs¹⁸ durch Vorschalten einer Metasuchmaschine¹⁹ generiert werden.

Auf der Grundlage von Überlegungen dieser Art hat eine Untergruppe der AG Verteilter Zeitschriftenserver ein Konzept für die Realisierung einer Verteilten Zugriffskontrolle auf der Basis der (Weiterentwicklung der) verteilten Authentifizierung und Autorisierung von ReDI-Konzepten der UB Freiburg in Kombination mit (Lizenz-) Daten und Diensten und des Rechtemanagements der EZB der UB Regensburg diskutiert und ausgearbeitet.



Abbild 5. Verteilte Zugriffskontrolle, Metasuche und Open Linking

¹⁵ Wir vermeiden hier die originäre von H. van de Sompel und P. Hochstenbach eingeführte Bezeichnung „SFX“ weil diese inzwischen für den Namen eines Produktes von Ex Libris verwendet wird, in dem die Firma das Ausgangsverfahren weiterentwickelt und vermarktet.

¹⁶ Document Object Identifier (DOI) der CrossRef-Foundation können, wie van de Sompel, Beit-Arie et al. gezeigt haben, in OpenURLs umgesetzt werden und umgekehrt, sind also mit diesen äquivalent.

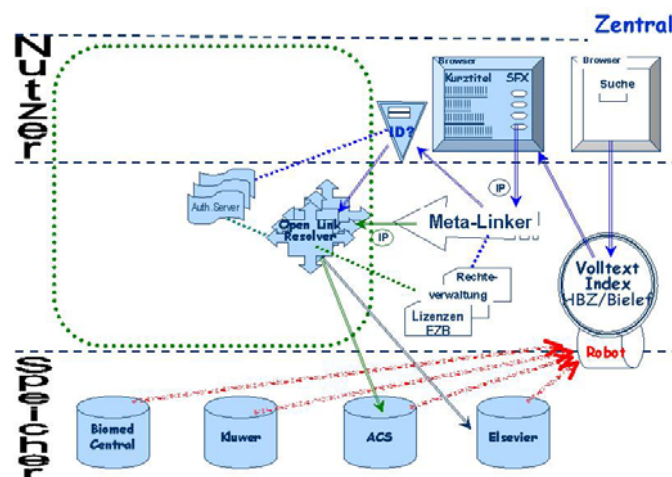
¹⁷ Wie etwa Preprint-Server der Fachgesellschaften oder allgemeine Dokumentenspeicher in Universitäten.

¹⁸ Falls das Query-Interface entsprechende Metadaten ausliefert.

¹⁹ Wie z. B. durch Vorschalten der MetaLib der Firma Ex Libris.

Falls ein Benutzer sich bei einer lokalen Bibliothek, die am VDS teilnimmt, etwa per ID und Passwort angemeldet hat, kann das Linking-Management (von der lokalen Bibliothek) benutzt werden, um den Benutzer auf den entsprechenden Server „durchzuschalten“ (vgl. das obige Schaubild).

Eine der Grundideen dieses Konzeptvorschlages besteht darin, dass Nutzer zunächst anonym im Zeitschriftenbestand suchen können und sich erst beim Zugriff auf ein Dokument identifizieren und damit „ihre“ Zugriffsberechtigung nachweisen müssen (deferred authentication). Diese Information wird von einer Komponente, die hier als Meta-Linker bezeichnet wird, beim Benutzer nachgefragt (s. das folgende Schaubild).



Abbild 6. Verteilte Zugriffskontrolle, Internet-Robot und Meta-Linker

In diesem Schaubild kann der Nutzer zunächst frei (im Internet) in dem von einem Internet-Robot erzeugten Volltext-Index des VDS suchen und gibt erst beim Click auf eine gefundene Titelanzeige den Ort „seiner“ Bibliothek und seine Identifikation preis. Der Meta-Linker kennt nun den entsprechenden lokalen Open-Linking-Server und kann den Benutzer zu diesem „durchschalten“.

Verifikation

Die KOBV-Zentrale setzt die Mechanismen des Open Linking inzwischen in seinem KOBV-Informationsportal (<http://digibib.kobv.de>) ein, dessen Freigabe für das Jahresende 2003 vorgesehen ist. Die Funktionsweisen der Verbindung von Metasuche und Open Linking sind nicht nur im KOBV, sondern auch in anderen Verbänden (national und international) bereits seit einiger Zeit erprobt. Die Universitätsbibliothek Bielefeld hat in ihrem Bericht zur Evaluation von Suchmaschinen-Technologie [5] aufgezeigt, dass eine vergleichbare Verbindung auch für die FAST-Suchmaschine eingerichtet werden kann. Weiteres dazu und zu FAST findet sich im folgenden Abschnitt.

Die Verifikation der Konzeptionen zur Verteilten Authentifizierung und Autorisierung in einer realistischen „Laborsituation“ lag jedoch außerhalb der Reichweite des Vorprojektes.

Phase 3 - Nahtlose Navigation und Suche

Metasuche und Open-Linking sind typische Komponenten eines lokalen oder regionalen Portals. In Deutschland sind Metasuchmaschinen z. T. weit verbreitet, in Berlin und Brandenburg auf der Basis der MetaLib von Ex Libris und Elektra von Sisis, in der DigiBib Nordrhein-Westfalens auf der Basis des IPS-Systems der Firma IHS. In Bayern wird der flächendeckende Einsatz der Systeme MetaLib und Elektra vorbereitet. Die Max-Planck-Gesellschaft setzt MetaLib als gemeinsames Portalsystem ihrer über ganz Deutschland verteilten Institute ein. Der HebIS-Verbund hat ein eigenes Portalsystem entwickelt und bereitet jetzt die Auswahl einer kommerziellen Portallösung vor. Auch der GBV und das BSZ bereiten sich auf den Einsatz einer Portalsoftware vor.

Portalsysteme auf der Basis von Metasuchmaschinen sind eng mit den lokalen Gegebenheiten und Bedingungen verknüpft. Ihre Stärke liegt in den Optionen der Personalisierung, mit denen Benutzer diese Art der verteilten Suche in gewissem Maße an ihre besonderen Bedingungen anpassen können. Mittels der Metasuche können sie die Suche, technisch handelt es sich um eine parallel verteilte Suche, auf eine Auswahl von Zielsystemen richten, auf OPACs oder auf bestimmte Datenbanken, soweit diese an die Metasuchmechanismen angeschlossen sind. Dadurch wird eine vom Nutzer gesteuerte Art der fachlichen Ausrichtung der Suche²⁰ möglich. Der Nachteil von Metasuchmaschinen liegt in bestimmten technischen Inkompatibilitäten und Ineffizienzen. Es macht wenig Sinn und bei manchen Metasuchmaschinen ist es auch technisch nicht möglich, alle an sie angeschlossenen Zielsysteme simultan abzusuchen. Eine Metasuche ist deshalb nicht die Methode der Wahl für das Suchen im Gesamtbeständen. Eine Metasuche lässt sich auch nur schwerlich in andere Portalsysteme integrieren. Kurz: für den VDS ist der Aufbau eines effizienten gemeinsamen Index notwendig.

Vorüberlegungen und Erfahrungen im KOBV haben zu dem Ergebnis geführt, dass es nicht opportun ist, den gemeinsamen Index des VDS auf der Basis eines Bibliothekssystems aufzubauen. Die These, für die Suche in allgemeinen Dokumentenbeständen aus dem Bereich der Wissenschaften einen Google-artigen Internet Robot einzusetzen, hatte das ZIB erstmalig in eine von K. Neubauer in Berlin einberufene Strategiediskussion deutscher Bibliothekare über „Die zukünftige Suchmaschine“ eingebracht. Die UB der Universität Bielefeld hat eine vergleichbare Idee für ihr Ziel des Aufbaus einer allgemeinen Wissenschaftssuchmaschine näher untersucht und sich für das in Norwegen entwickelte FAST-System der Firma Fast Search&Transfer entschieden. Google schied schon aus formalen Gründen für diesen Zweck aus. Die entsprechende Software ist nur als geschlossenes System für ein Intranet erhältlich. Modifikationen an der Google-Software seitens der UB Bielefeld wären damit nicht möglich.

Die Frage, ob sich das FAST-System auch als Internet-Robot und Suchmaschine für den VDS eignet, hat die UB der Universität Bielefeld im Rahmen eines Auftrags des ZIB als Teil des vorliegenden Vorprojektes genauer untersucht [5] und befürwortet. Insbesondere waren durch diese Untersuchung auch die folgenden Fragen zu klären:

- Ist das FAST-System leistungsfähig genug, um einen sehr großen Bestand an digitalen Volltexten automatisch zu sammeln und zu indexieren?
- Ist es technisch im lokalen Netz und auch über das Internet skalierbar?

²⁰ Allerdings nur insoweit, wie die angeschlossenen Systeme selbst fachlich hinreichend separiert sind. Das ist bei OPACs typischerweise nicht der Fall, selbst dann nicht, wie es in gewissen Fällen bei OPACs von Sonder-sammelgebetsbibliotheken der Fall ist, wenn ein Klassifikationsschema eingesetzt wird. Der OPAC der SUB-Göttingen enthält z.B. nicht nur Literatur der Mathematik. Integriert man diesen OPAC in eine Meta-Suche, so wird die Suche im Allgemeinen nicht nur mathematische Literaturverweise liefern.

- Wäre auf der Basis von FAST eine kooperative Indexierung durch das HBZ, den KOBV und die UB der Universität Bielefeld möglich?
- Bietet das FAST-System eine Schnittstelle derart, dass die Erzeugung von OpenURLs in die Software integriert werden kann?
- Verfügt es über die notwendigen Fähigkeiten zur Konfiguration entsprechender Ranking-Verfahren? Wäre ein Google-artiges Link-basiertes Ranking möglich?
- Ist das System einfach zu handhaben?
- Beinhaltet der FAST-Robot Automatismen zum Sammeln von Information über Webseiten und via ftp aus den File-Systemen?

Die Ergebnisse dieser Evaluation sind für die Suchmaschine FAST positiv verlaufen und sind diesem Bericht als Anlage beigefügt. Sie waren die Grundlage dafür, dass sich das ZIB und die KOBV-Zentrale für den Aufbau einer Internet-Suchmaschine (Hardware und Software) für den VDS des KOBV auf der Basis von FAST entschieden hat und ebenfalls evaluiert, ob FAST auch eine geeignete Grundlage für die Realisierung des VDS ist (zu Ergebnissen siehe weiter unten).

Abschließend sei angemerkt, dass es nur einen gemeinsamen Index im VDS geben soll, der aber in die lokalen Portalsysteme der Verbünde und Bibliotheken integriert werden kann. Er steht dann aber auch als eigenständiger Suchmechanismus im Web zur Verfügung. Im Gegensatz dazu soll es mehrere Metasuchmaschinen geben, die auf die Spiegel und Speicher des VDS verweisen, eine je Verbundsystem, und diese sind möglicherweise voneinander unterschiedlich.

Es ist zu erwarten, dass auch eine größere Anzahl von lokalen Bibliotheken eigene Portalsysteme einsetzen. Auch Link Resolver werden in größerer Zahl und heterogener Ausprägung im VDS auftreten. Die Architektur des Gesamtsystems ist so geartet, dass weitere Bibliotheken mit eigenen Speichern relativ einfach hinzutreten und auch mit ihren Portalsystemen von den Leistungen des VDS partizipieren können. Es ist auch durchaus denkbar, dass sich die Teilnehmer aus dem Kreis von vascoda, insbesondere die Informationsverbünde mit eigenen Spiegeln und Speichern am VDS beteiligen. Rein technisch gesehen steht dem nichts im Wege.

Verifikation

Die Möglichkeiten, Fähigkeiten und auch die Schwächen der Suchmaschine FAST bei der Indexierung eines substantiellen Datenbestandes (das Testbed ist durch die in [3] beschriebenen Kluwer-Daten des FAK definiert) sind in einem gesonderten Erfahrungsbericht zusammengestellt [4]. Die Aufgabe dieses Projektteiles bestand darin, die Eignung von FAST für die Realisierung des Verteilten Dokumentenspeicher des KOBV systematisch und gründlich zu analysieren und zu überprüfen. Als Testmodell sollte die in Abb. 1 skizzierte vereinfachte Modellvorstellung dienen. Konkret handelt es sich um einen im Internet verteilten Speicher. Zu den Kluwer-Daten, die im ZIB aufliegen, sollte ein kleinerer Bestand von BioMed-Central-Daten, die an der Universitätsbibliothek Potsdam aufliegen, hinzutreten und ebenfalls mittels FAST indexiert werden. Beide Datenbestände stellen auch Metadaten bereit. Die Analyse sollte deshalb auch die Frage beantworten, ob die in Abbild 1 skizzierten Meta-Indexe ebenfalls durch FAST erzeugt werden können.

Beim Einsatz von FAST mit „Massendaten“ aus dem Verlagsbereich zeigt sich, dass für die Indexierung der Volltexte mit einem substantiellen Aufwand für die Dokumentenvorbereitung

gerechnet werden muss. FAST enthält einen Internet-Robot, der auf HTML- und XML-Seiten spezialisiert ist. Verlage liefern hingegen i. d. R. PDF-Dateien (unterschiedlichster Ausprägungen) und Metadaten in einem XML- oder SGML-Format, die entsprechend umzuwandeln sind. Dafür stellt FAST zwar Werkzeuge bereit, beim Einsatz in einer realistischen Betriebsumgebung stellt sich jedoch heraus, dass diese oft nicht ausreichen, dass eigene Programmierung erforderlich ist und dass (nicht nur in diesem Falle) die Handhabung von FAST systemtechnische Erfahrungen voraussetzt. Es ist kein Werkzeug für den Bibliothekar im Sinne eines integrierten Bibliotheksystems, sondern eher ein Werkzeugkasten für den Systemspezialisten mit Informatik-Kenntnissen. Das ist sicherlich kein prinzipielles Problem, jedoch sind ohne solche Kenntnisse die Probleme des Ladens der Daten²¹ kaum lösbar.

Unabhängig davon lassen sich, sofern die Datenvorverarbeitung auf den lokalen Speicher verlegt und die Daten dort dem FAST-System in ihrer konvertierten Form durch einen internen Web-Server (etwa Apache) zugänglich gemacht werden, aus den bisherigen Erfahrungen mit FAST die folgenden Schlüsse ziehen:

- Der verteilten Volltext-Indexierung eines substantiellen (Kluwer) Datenbestandes über das Internet stehen keine prinzipiellen Probleme entgegen.
- Die Laufzeiten halten sich auf einem Sun-Solaris-System (3 GB RAM, 1 Prozessor zu 900 MHz, mit 300 GB Festplatte) in vertretbaren Grenzen. Besser wäre der Einsatz eines Hochleistungs-PCs (Linux, Prozessoren zu 3 GHz, 1 Terabyte Festplatte).
- Ein (lokaler) Metadaten-Index lässt sich mit FAST aus vorhandenen Metadaten erstellen und auch ein im Internet verteilter Metadaten-Index.
- Falls keine Metadaten zur Verfügung stehen, muss auf einen Metadaten-Index verzichtet werden. Die Extraktion von Metadaten aus PDF-Files ist ein F&E-Projekt.
- Die Volltextindexierung des BioMed-Central-Spiegels, dessen Dokumente bereits als HTML-Text im Web zur Verfügung stehen, ist eine kleinere Aufgabe für FAST.
- Open URLs müssen im Zuge der Datenvorverarbeitung erzeugt werden. Der lokale Speicher ist so zu konfigurieren, dass er als Target für Link-Resolver dienen kann.

Der für eine Anbindung an eine Metasuchmaschine kritische Fall ist dann gegeben, wenn weder vom Verlag noch von einer anderen Quelle Metadaten zu Artikeln erhältlich sind. An die Stelle der Suche nach Metadaten muss dann ersatzweise die Suche in einem lokalen Volltextindex treten, der die Rolle eines Basic-Index übernimmt.

Die Erfahrungen bei der Indexierung mit der Suchmaschine FAST machen deutlich, dass die Verarbeitung von Massendaten auch im Kontext der Anwendung einer Internet-Suchmaschine immer mit einem erheblichen Aufwand verbunden ist. Wäre das nicht der Fall (und würden durch den Einsatz eines Internet-Robots alle Probleme dieser Art verschwinden), so würde keine Notwendigkeit dazu bestehen, die Arbeit auf „viele Schultern“ (Verbünde und Bibliotheken) zu verteilen. Genau das ist aber ein zentraler Gedanke des VDS.

Der Erfahrungsbericht zum Einsatz der Suchmaschine FAST ist zum Zeitpunkt der Abfassung dieses Schlussberichtes noch nicht abgeschlossen. Die Probleme und Kosten des sicheren und fehlerfreien Transfers und Updates großer Datenmengen über das Internet wurden z. B. bisher nicht analysiert. Eine genaue Aufwands- und kostenorientierte Betrachtung steht ebenfalls noch aus. Der Erfahrungsbericht bedarf ferner noch der vertraglich zugesicherten Zustimmung durch die Herstellerfirma, die bezüglich des Verbots der Publikation systeminterner Eigenschaften, die als vertraulich zu behandeln sind, ein Einspruchsrecht hat.

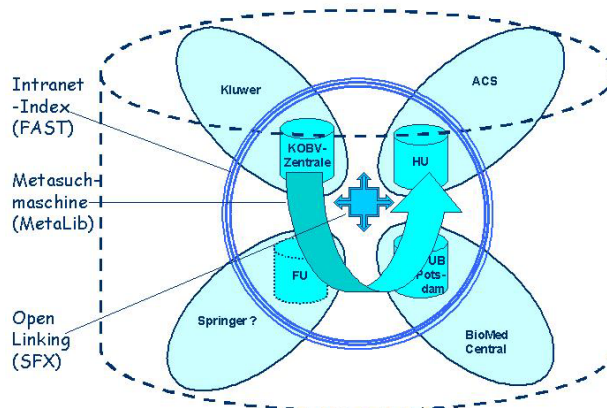
²¹ Elsevier liefert für den Einsatz seiner ScienceDirect-Software auf Elsevier-Content abgestimmte Lader mit.

Phase 4 – Verteilter Betrieb, Dauerhaftigkeit

Für die Sicherung der langfristigen Verfügbarkeit gibt es keine Patentlösung, allenfalls gewisse Strategien, die eine äußerst disziplinierte Vorgehensweise voraussetzen und selbst dann das gewünschte Ziel allein nicht garantieren können. Erfolg versprechende Ansätze²² sind:

- I. Verwendung möglichst elementarer Bausteine/Softwaremodule in der Weise, dass eine komponentenweise Modernisierung und rechtzeitige Migration möglich ist
- II. Einsatz Offener Schnittstellen und international verbreiteter Protokolle
- III. Einsatz von Internet-Technologien, Web-Schnittstellen

Der „Verteilte Ansatz“ des KOBV²³, der auch dem VDS zugrunde liegt, verfolgt alle diese Strategien. Der bisher übliche Verbundkatalog wird im KOBV als Verteilter Katalog („bestehend“ aus der Sammlung lokaler OPACs) aufgefasst, der mittels der KOBV-Metasuchmaschinen virtuell und mittels des Gemeinsamen KOBV-Index „physisch“ zusammengeführt wird. Das „Bindemittel“ sind offene, internationale Schnittstellen und Protokolle. Einer der wesentlichen Vorteile der „Zerlegung“ in einzelne verteilte Module besteht darin, dass nun die jeweiligen Bausteine von den lokalen Bibliotheken ausgewechselt und damit schrittweise modernisiert und auf den neuesten Stand der Technik gebracht werden können.



Abbild 8. Virtueller Dokumentenspeicher des KOBV

In vergleichbarer Weise lässt sich auch ein Dokumentenspeicher als aus elementaren Komponenten zusammengesetzt auffassen, der nur virtuell existiert und durch Mechanismen der verteilten Indexierung adressierbar²⁴ wird. Das ist der Ansatz des VDS des KOBV. Die „Überlebenschancen“ der einzelnen Komponenten können als um so besser eingeschätzt werden, je elementarer sie sind und je besser sie sich an Internet und Web-Technologien²⁵ orientieren. Es

²² Die an dieser Stelle immer wieder genannte Emulationsstrategie soll in diesem Zusammenhang nicht betrachtet werden, weil Zeitschriften und Dokumente miteinander und im Web vernetzt sind (oft sogar dynamisch) und sich „die Außenwelt“ (das Web) permanent ändert. Dieser Sachverhalt ist nicht emulierbar.

²³ Erstmals realisiert in der KOBV-Suchmaschine, die eine Metasuche und einen gemeinsamen Index umfasst.

²⁴ Technisch bedeutet dieses, dass es in den lokalen Speichern jedem Dokument eine eindeutige System-ID gibt, über die das einzelne Dokument ansprechbar und wartbar ist.

²⁵ Diese Aussage scheint paradox zu sein, da Internet und Web in ständiger Bewegung sind und sich dauernd wandeln. Tatsächlich sind aber die technischen Grundkomponenten von Internet und Web von großer Dauerhaftigkeit, wie es bestimmte Internet-Protokolle (ftp, Email) und Web-Dienste (Apache-Server) demonstrieren.

liegt nahe, das generische Modell des VDS auch im VDS einzusetzen²⁶. Mit anderen Worten: Der Verteilte Dokumentenspeicher tritt im VDS als eigenständiger Objektspeicher auf. Es handelt sich aber um eine über das Internet verteilte Systemarchitektur.

Die einzelnen Objektspeicher des VDS weisen alle technischen Funktionalitäten auf, über die Spiegelserver von Verlagseinhalten verfügen. Sie speichern und stellen z.B. bereit:

- Hierarchisch gegliederte Sammlungen digitaler Journale
- Volltexte, Table of Contents, Indexe (von PDF-Dateien bis zu LaTeX-Texten)
- Navigationsstrukturen zum Browsen und Suchen (oft einander durchdringend)
- Multimediadokumente, bis hin zu neuen Publikationsformen auf Web-Servern
- Metadaten zur Suche nach Attributen (gelegentlich nach dem Dublin Core)
- Module für Authentifizierung und Autorisierung, Rechtemanagement
- Web-Interfaces auf XML-Basis, Internet- und Web-Protokolle
- Nutzungsstatistiken in diversen Formaten, ggf. verlagsbezogene Formen

Zusätzlich erfüllen Objektspeicher spezifische Anforderungen, die sich aus der Funktionalität und der verteilten Organisationsform des VDS - und in der sich anschließenden Phase für den VZS - ergeben, wie z.B.

- Direkte Ansteuerung von Dokumenten mittels URI²⁷/URN²⁸/SID²⁹
- Eignung als Ziel von Open URLs³⁰
- OpenURL-Fähigkeit³¹
- Autorisierung und Rechtemanagement
- Eignung für die Speicherung von Sammlungen/Metadaten/Navigationsstrukturen digitaler Journale und Dokumente unterschiedlicher Anbieter

Außerdem müssen Objektspeicher über Schnittstelleneigenschaften verfügen, wie z.B.

- Zum Export und Import von Dokumenten, Metadaten und Navigationsstrukturen
- Für die technische Sicherung von Daten (z.B. auf Bandroboter)
- Zur Datenbank/zum Filesystem, auf dem sie aufsetzen

Von daher sind Objektspeicher hochspezialisierte Systeme, die – sofern als integrierte Systeme ausgelegt – mit ganz bestimmten Risiken verbunden sind. Der KOBV hat einen DFG-Antrag gestellt, um Fördermittel für die Umsetzung dieser Idee einzuwerben.

²⁶ Genauer: die Verwendung zu erproben. Es handelt sich um ein mit Risiken behaftetes Entwicklungsprojekt.

²⁷ Uniform Resource Identifier

²⁸ Uniform Resource Name

²⁹ System Identifier

³⁰ genauer: von URLs, die Open-Link-Resolver aus der OpenURL erzeugt.

³¹ Nicht unbedingt notwendig, da die OpenURL durch Vorschaltung gewisser Metasuchmaschinen zu Suchergebnissen in Servern, die geeignete Metadaten anbieten, automatisch erzeugt werden kann.

Referenzen

- [1] Status: Verteilter Dokumentenserver (VDS/VZS); Stand: 14.09.03
Anlage zum Protokoll der außerordentlichen Sitzung des vascoda-
Steuerungsgremiums und der AG-Verbundsysteme am 16.09.03, Hannover
- [2] Rahmenantrag zur Förderung durch BMBF und DFG – Entwurf, Stand 05.12.03
,Aufbau eines „Verteilten Dokumentenservers“ (VDS) zur Schaffung einer
offenen digitalen Bibliothek für die Wissenschaft’ (Vorhabenbeschreibung)
- [3] Thorsten Koch; Verteilter Dokumentenspeicher –
Erfahrungen mit den Kluwer Daten des Friedrich-Althoff-Konsortiums
Konrad-Zuse-Zentrum für Informationstechnik Berlin, 16. Dezember 2003
- [4] Zara Kanaeva; Verteilter Dokumentenspeicher –
Erfahrungen mit der Suchmaschine FAST; Entwurf, Stand 16. Dezember 2003
Kooperativer Bibliotheksverbund Berlin-Brandenburg, KOBV-Zentrale im ZIB
- [5] Carola Kummert, Friedrich Summann, Sebastian Wolf
Abschlussbericht zur Evaluation von Suchmaschinen-Technologien
im Vorhaben „Verteilter Zeitschriftenserver“ (VZS)
Universitätsbibliothek Bielefeld, Januar 2003
- [6] Virtueller Dokumentenspeicher des KOBV
DFG-Projektantrag vom 18. September 2003 (nur intern verfügbar)
Kooperativer Bibliotheksverbund Berlin-Brandenburg
- [7] Joachim Lügger, Michael Rieck, Michael Voß
Konzeptvorschlag für die Speicherung digitaler Dokumente im KOBV
Beschluss des KOBV-Hauptausschuss, 4. April 2002
- [8] Konzeptvorschlag für die Speicherung digitaler Dokumente im KOBV
Beschluss des Kuratoriums des KOBV vom 7. Mai 2002
- [9] Verteilter Zeitschriften-Server (Rahmenplan)
Skizze eines Vorhabens der AG der Verbundsysteme
AG Verteilter Zeitschriftenserver der AGVerbund, 26.09.02, DDB, Frankfurt

Anhang

Bemerkungen zur Nachhaltigkeit

Die Arbeitsgemeinschaft der Verbundsysteme hat sich von Beginn an intensiv am Vorhaben VZS und am VDS (auf der Ebene der Verbünde) sowie an den diesbezüglichen Vorarbeiten und Spezifikationen beteiligt. Anders als das ZIB/KOBV erhielten die Verbundsysteme hierfür keine finanzielle Unterstützung. Auch die an dieser Initiative beteiligten Entwicklungszentren erhielten mit Ausnahme der Universitätsbibliothek Bielefeld weder Personal- noch Sachmittel für Ihre Beteiligung. Der KOBV hat das Vorprojekt nach seinem formalen Schluss am 31.03.03 aus eigenem Interesse und aus eigener Kraft mit erheblichen Eigenanteilen personeller und sachlicher Art fortgeführt. Das hohe Interesse, das die Verbünde und ihre Partner an dem Aufbau und dem Betrieb des VDS, wie er nun im vascoda-Kontext bezeichnet wird, haben, drückt sich in den eigenen Leistungen und Beiträgen an diesen gemeinsamen Aktivitäten³² aus. Dieses soll in den folgenden Abschnitten dokumentiert werden.

Initiativen, Workshops und Arbeitstreffen

Dieser Abschnitt nennt und kommentiert in knapper Form die Arbeitsthemen und –treffen der einzelnen Untergruppen der Arbeitsgruppe Verteilter Zeitschriftenserver, der Vollständigkeit halber einschließlich der Termine vor dem Start des Vorprojektes „Offene Navigation und Suchmaschinen“. Die einzelnen Gruppen orientierten sich an den Abstimmungen und Vorgaben aus dem 1. Treffen der AG Verteilter Zeitschriftenserver am 06.06.2002 am ZIB in Berlin und dem 2. Treffen am 26.09.2002 in der Deutschen Bibliothek in Frankfurt (Rahmenplan). Erste Zwischenergebnisse wurden der Arbeitsgemeinschaft der Verbundsysteme in der Sitzung am 22.10.2003 vorgestellt, die ebenfalls in der Deutschen Bibliothek in Frankfurt stattfand. Die GASCO war auf dem Wege über ihre Vertreter Herrn M. Rieck (FAK) und später Frau Dr. H. Schaeffler (BSB), in der AG Verteilter Zeitschriftenserver ständig eingeschaltet und informiert³³.

Die folgende Zusammenstellung gibt eine Übersicht über die Arbeitstreffen der AG Verteilter Zeitschriftenserver und ihrer Arbeitsgruppen in zeitlicher Reihenfolge (keine Wertung und ohne KOBV-interne Arbeitstreffen):

Januar 2002	KOBV Berlin	KOBV-Hauptausschuss: „Konzeptvorschlag zur Speicherung digitaler Dokumente im KOBV“
20.02.2002	HBZ Köln	Beschluss der AGVerbund (Einrichtung AG VZS)
07.05.2002	ZIB Berlin	Beschluss des KOBV-Kuratoriums zum VDS
06.06.2002	ZIB Berlin	1. Treffen der AG VZS (Konstituierung)
16.06.2002	TIB Hannover	Erstkontakt mit VLib-Geschäftsstelle
18.07.2002	UB Regensburg	VZS-Redaktion und Integration der EZB
05.08.2002	BSB München	Rolle der BSB, Digitalisierung
07.08.2002	Univ. Hamburg	Object Store
26.09.2002	DDB Frankfurt	2. Treffen der AG VZS (VZS-Rahmenplan)
16.10.2002	UB Bielefeld	Intranet-Index, FAST-Suchmaschine
18.10.2002	UB Freiburg	Verteilte Authentifizierung und Open Linking
22.10.2002	DDB Frankfurt	AGVerbund: VZS-Bericht und Vorgehensweise
07.11.2002	Univ. Essen	Objekt-Speicher und MyCoRe

³² Eine ausführliche Dokumentation der Treffen findet sich auf dem internen Server <http://kirke.hbz-nrw.de/vzs/>

³³ M. Rieck ist als ständiger Gast an den Beratungen zum Verteilten Dokumentenspeicher des KOBV beteiligt.

22.11.2003	Univ. Bielefeld	FAST-Workshop
03.12.2002	LB Darmstadt	Meta-Suche, Meta-Linker und Zugriffskontrolle
17.12.2002	SBB Berlin	Rolle der ZDB im VZS
20.12.2002	BSB München	Koordination mit GASCO
17.01.2003	ZIB Berlin	Workshop: Internet-Suchmaschinen (FAST)
23.01.2003	Univ. Frankfurt	3. Treffen der AG VZS (Projektskizzen)
25.02.2003	DDB Frankfurt	Langzeitverfügbarkeit, ZDB-Integration
05.03.2003	DDB Frankfurt	GASCO: VZS-Bericht und Koordination
06./07.03.03	DDB Frankfurt	Treffen vascoda und Verbundsysteme
26.03.2003	ZIB Berlin	AGVerbund: VZS-Bericht und Projektstruktur
05.06.2003	DDB Frankfurt	4. Treffen der AG VZS (Content der Verbünde)
31.07.2003	BSB München	Verteilte Zugriffskontrolle und Open Linking
07.08.2003	ICC Berlin	Rolle und Integration von EZB und ZDB
12.09.2003	BZS Konstanz	Metadatenverwaltung und Metadaten-Registry
16.09.2003	TIB Hannover	vascoda und AGVerbund/VDS

Die AG VZS besteht aus folgenden Mitgliedern (Stand September 2003)

BSB	J. Kunz	Dr. H. Schaeffler (GASCO)
BVB	Dr. F. Geißelmann	Dr. E. Hutzler (EZB)
DDB	R. Altenhöner	H. Liegmann (Langzeitverfügbarkeit)
GBV	R. Diedrichs	[Dr. St. Gradmann (GAP)]
HBZ	Dr. P. Kostädt	F. Summann (Suchmaschinen)
Hebis	M. Michalek	Dr. M. Bischoff (Metabroker)
KOBV	M. Kuberek	Dr. M. Voß (LDAP)
SBB	B. Sigrist	U. Schulz (ZDB)
SWB	V. Conrad	B. Oberknapp (ReDI)
Koordination	J. Lügger	

Jedes Verbundsystem ist in der Arbeitsgruppe durch ein von der Verbundzentrale benanntes Mitglied (mittlere Spalte) vertreten und durch ein in Abstimmung mit der betreffenden Verbundzentrale vom Koordinator der Arbeitsgruppe aus fachlichen Gründen einberufenes Mitglied (rechte Spalte). Auf Grund ihrer spezifischen Erfahrungen wurden B. Oberknapp und A. Ruppert (ReDI) von der UB Freiburg sowie F. Seiffert (HBZ) und S. Litsche (KOBV) zu einzelnen Beratungen hinzugezogen.

Die AG Verteilter Zeitschriftenserver hatte sich bereits bei ihrem ersten Treffen am 06.06.2002 in Berlin darauf verständigt, sich in ihrer Zielsetzung auf die technischen Probleme zu beschränken, die bei der Speicherung lizenzierter Journale zu lösen sind. Dem generischen Modell des KOBV hat sie die folgenden technischen Elemente konzeptionell hinzugefügt:

- Metadatenverwaltung, Metadaten-Registry
- Verteilte Authentifizierung als Kombination von
 - Verteilter Authentifizierung und Autorisierung
 - Rechtemanagement
 - Meta-Linker und nahtlose Navigation
- Virtueller Objekt- und Dokumentenspeicher
- Einfaches Browsing auf der Grundlage von Metadaten der EZB
- Interface zum Depotsystem der DDB

Bis zum Jahresende 2002 entstanden dazu ausgereifte Projektskizzen und im Zuge des Zusammengehens mit vascoda wurden diese im Verlauf des 4. Quartals 2003 zu vollständigen Projektanträgen und einem VDS-Gesamtkonzept ausgearbeitet.