

Open Source Discovery System VuFind an sächsischen Hochschulbibliotheken

Jens Lazarus

Ausgangslage: Verändertes Nutzerverhalten, verändertes Angebot

Mit dem Einsatz von Discovery Systemen reagieren Bibliotheken auf veränderte Nutzererwartungen¹ und auf ein massiv verändertes Medienangebot.

Kataloge von Bibliotheken haben die Entwicklungen von Internetanwendungen und Rechercheplattformen bislang nur eingeschränkt nachvollzogen. Das Rechercheverhalten von Nutzerinnen und Nutzern wird von Google und anderen Plattformen geprägt. Dagegen fallen die bibliothekarischen Online-Kataloge weit zurück. Die Funktion von Bibliothekskatalogen wird nutzerseitig häufig reduziert auf den Besitz- und Standortnachweis für an anderer Stelle ermittelte Informationsquellen. Aufwändige bibliothekarische Dienstleistungen wie die Sacherschließung und andere Teilbereiche der Katalogisierung laufen ins Leere.

Im Vergleich zum gedruckten Bestand sind elektronische Medien geprägt von einer hohen Dynamik, von komplexen Lizenzierungsmodellen und einer sehr unterschiedlichen Situation in Bezug auf die Verfügbarkeit von Metadaten. In den ursprünglich für den Nachweis gedruckter Medien konzipierten Bibliothekssystemen lassen sich elektronische Medien massenhaft nur bedingt abbilden. So werden diese in separaten Nachweissystemen geführt und finden nur bedingt Eingang in den Bibliothekskatalog. Für die Recherche wechseln Nutzerinnen und Nutzer in unterschiedliche Datenbanken.

Mit verschiedenen technischen Lösungen und unterschiedlichen bibliothekarischen Konzepten gehen Bibliotheken seit einigen Jahren diese Problematik an. Ziel ist es, die Ressourcen einer Bibliothek unter einer Oberfläche komfortabel und qualifiziert zugänglich zu machen. Aus der Riege der Bibliotheken im universitären Kontext können beispielhaft die SLUB Dresden, USB Köln, UB Konstanz oder die UB Mannheim genannt werden.

Mit dem Einsatz von Discovery Systemen reagieren Bibliotheken auf veränderte Nutzererwartungen und auf ein massiv verändertes Medienangebot. In einem Konsortium von Universitäts- und Hochschulbibliotheken in Sachsen kommt das Open Source Discovery System VuFind unter Anbindung des aggregierten Index Primo Central der Firma Ex Libris zum Einsatz. Der Artikel beschreibt das Projekt finc und die technischen Grundlagen dieser Lösung.

Libraries respond to changing user expectations and the shift to electronic publication in media market by the application of Discovery Systems. In a consortium of university and college libraries in Saxony, the Open Source Discovery System VuFind is used in connection with the aggregated index Primo Central by Ex Libris. The article describes the project named finc and the technical approach taken.

Konsortium, Zielstellung

Mit dieser Ausgangslage fand sich im Jahr 2010 eine Interessengemeinschaft der folgenden sächsischen Universitäts- und Hochschulbibliotheken unter der Leitung der Universitätsbibliothek Leipzig zusammen.

1. Universität Leipzig
2. Technische Universität Chemnitz
3. Technische Universität Bergakademie Freiberg
4. Hochschule für Technik und Wirtschaft Dresden
5. Hochschule für Technik, Wirtschaft und Kultur Leipzig
6. Hochschule Zittau/Görlitz
7. Westsächsische Hochschule Zwickau
8. Hochschule für Bildende Künste Dresden
9. Hochschule für Musik Carl Maria von Weber, Dresden
10. Hochschule für Musik und Theater „Felix Mendelssohn Bartholdy“ Leipzig
11. Hochschule für Grafik und Buchkunst Leipzig

Zielstellung für die beteiligten Bibliotheken war es, den bislang angewandten WebOPAC des Lokalsystems Libero durch eine suchmaschinenbasierte Rechercheoberfläche in Verbindung mit einem aggregierten Index in konsortialer Anwendung zu ersetzen.

¹ <http://www.jisc.ac.uk/media/documents/publications/reports/2010/digitalinformationseekerreport.pdf> [28. Mai 2012]

Marktsichtung, Evaluierung

Die Suche nach geeigneten Produkten und Lösungen führte zu einer ausführlichen Sichtung am Markt vorhandener kommerzieller Produkte. Gleichzeitig beschäftigte sich ein Team der Universitätsbibliothek Leipzig und der Hochschule für Musik und Theater Leipzig mit verfügbaren Open Source Lösungen. Um hier die notwendige Kompetenz aufzubauen und zu einer fundierten Einschätzung zu kommen, wurde eine Testinstallation mit dem Open Source System VuFind realisiert. Innerhalb weniger Wochen konnte für diese Aufgabenstellung eine überzeugende Machbarkeitsstudie (Proof of Concept) präsentiert werden.

Im Ergebnis der Evaluation wurde festgestellt, dass der zunächst angenommene Entwicklungsvorsprung der kommerziellen Systeme gegenüber dem getesteten Open Source System VuFind in den relevanten Funktionalitäten nicht bestätigt werden konnte. Ein

aus in Betracht gezogenen möglichen Nachteile einer Community-abhängigen, freien Software auf.

VuFind

Basis für die Entwicklung neuer Rechercheoberflächen für sächsische Hochschulbibliotheken ist VuFind², eine Open Source Plattform zur Gestaltung eigener suchmaschinenbasierter Discovery Services auf der technischen Basis von Apache Solr. VuFind hat ihren Ursprung an der Villanova University (Pennsylvania, USA) und wird inzwischen von einer wachsenden Community entwickelt und gepflegt. Zu den größten Anwendern von VuFind gehören die National Library of Australia, die National Library of Ireland, der Hathi Trust, die Yale University und eine Reihe weiterer Universitätsbibliotheken weltweit.

In Deutschland ist VuFind bislang unter anderem im GBV für die Nationallizenz-Recherche „Suchkiste“,

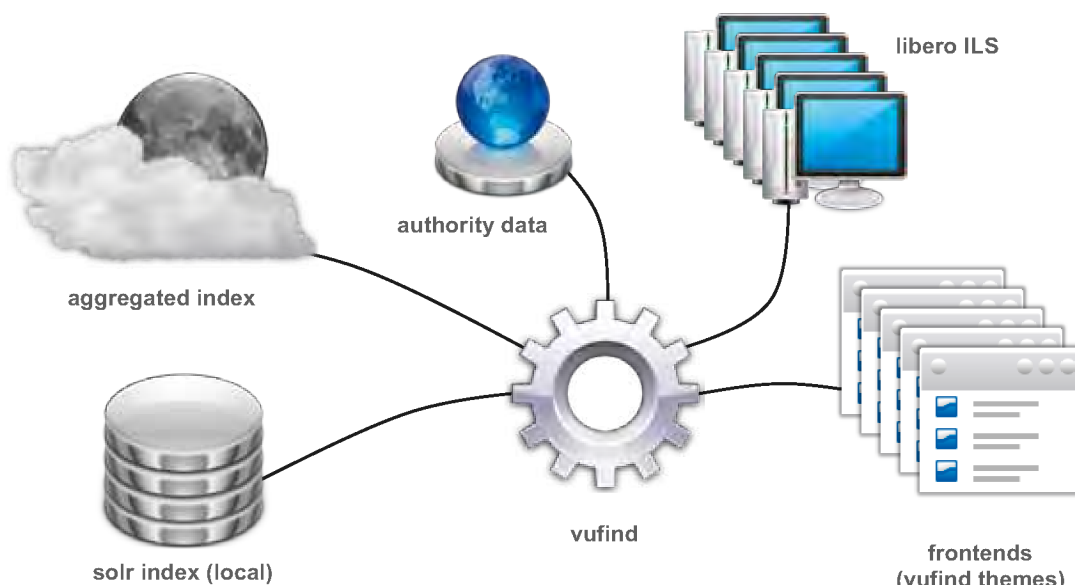


Abbildung 1:
VuFind im
Projekt *finc*

Abgleich wirtschaftlicher Rahmenbedingungen, wie Implementierungsaufwand, langfristiger Personaleinsatz oder Lizenzkosten, sah beide Lösungsansätze – kommerzielles Produkt und Open Source – ebenfalls nahe beieinander, so dass auf diese Weise im Rahmen des Konsortiums keine deutliche Präferenz ermittelt werden konnte.

Die Entscheidung der Interessengemeinschaft für den Einsatz der quelloffenen Software VuFind war vor allem eine strategische Entscheidung. Die Rechercheoberfläche einer Bibliothek ist das zentrale Werkzeug, mit dem eigene und externe Ressourcen angeboten und vermittelt werden. Für die beteiligten Einrichtungen wogen die hohen Freiheitsgrade, die Möglichkeit der selbstbestimmten Entwicklung, die Flexibilität in der künftigen Einbindung externer Dienstleistungen und Informationsquellen die durch-

für die Bielefeld Academic Search Engine (BASE), in der UB der TU Hamburg-Harburg und in der Zentralbibliothek des Forschungszentrums Jülich im Einsatz. Weitere Einrichtungen beschäftigen sich momentan intensiv mit den Möglichkeiten von VuFind. Das Interesse und die aktive Beteiligung an einem ersten deutschen Anwendertreffen am 26. und 27. Juni 2012 an der Universitätsbibliothek Leipzig war dementsprechend groß³.

Aggregierter Index

Neben den erweiterten Funktionalitäten des Discovery Systems sollte für die Bibliotheken der sächsischen Hochschulen eine grundsätzlich neue Quali-

2 <http://vufind.org/> [28. Mai 2012]

3 <http://blog.finc.info/vufind-session-2012> [28. Mai 2012]

tät in der Recherche über die Einbindung eines aggregierten Index erreicht werden. In einem solchen Index werden Datensätze aus einer Vielzahl von Datenbanken verschiedener Anbieter und Verlage zusammengeführt. Die von den Einrichtungen lizenzierten Zeitschriften werden auf Artekelebene oder in der Volltextsuche über eine Rechercheoberfläche erschlossen. Außerdem kann die Einbindung eines aggregierten Index auch die Recherche über verfügbare, externe Datenbestände, wie etwa die Digitale Bibliothek des Hathi hTrusts ermöglichen. Ein aggregierter Index kann mehrere hundert Millionen Datensätze umfassen.

Andere Datenangebote, wie etwa die Zeitschriften-Aufsatztiteldaten der Firma Swets (Swets-Scan Online-Content-Daten), wurden als technisch machbare Alternative geprüft. Allerdings lag hier die Abdeckung des aktuellen Zeitschriftenbestandes der Universitätsbibliothek Leipzig mit etwa 50% doch weit unter der nahezu vollständigen Erschließung der auf dem Markt befindlichen aggregierten Indizes.

Nach einer Ausschreibung, einem intensiven Evaluierungsprozess auf der Basis eines umfangreichen Kriterienkataloges⁴ und einer Teststellung in realer Umgebung fiel die Entscheidung für den aggregierten Index Primo Central der Firma Ex Libris⁵. Beurteilt wurden technische Aspekte, sowie organisatorische, funktionale und inhaltliche Leistungsmerkmale. Ausschlaggebend für die Entscheidung für Primo Central war vor allem die überzeugende technische Lösung zur Anbindung des Index an das Open Source Discovery System VuFind.

Realisierung, Projekt *finc*

Zur Implementierung des Open Source Discovery Systems und der Lizenzierung des aggregierten Index Primo Central wurden von der Universität Leipzig Fördermittel des Europäischen Fonds für regionale Entwicklung beantragt, die nach entsprechender Begutachtung im November 2011 bewilligt wurden. Der Freistaat Sachsen und die Europäische Union unterstützen das Projekt über eine Laufzeit von drei Jahren. Ab September 2011 konnten so eine Projektmitarbeiterin für den bibliothekarischen Bereich und vier Projektmitarbeiter im IT-Bereich eingestellt werden. Das Projekt erhielt den Namen *finc* und ist an der Universitätsbibliothek Leipzig angesiedelt. *finc* ist ein Kunstwort, das als *find in catalog* gelesen werden kann.

Um die Beteiligung und die nahtlose Kommunikation aller involvierten Bibliotheken zu sichern und für Transparenz über Entwicklungsfortschritte und anstehende Aufgaben zu sorgen, wird das Management-Tool Redmine⁶ eingesetzt. Das gesamte Projekt wird von einer Reihe engmaschiger Treffen und Workshops in der Region begleitet. Als erste in der Reihe der Konsortialteilnehmer ging im März 2012 die Universitätsbibliothek Leipzig mit einem neuen, suchmaschinenbasierten Katalog⁷ in der Beta-Version online.

Technische Grundlagen

Das Hosting der Suchmaschinen, Datenbanken und Webserver für alle beteiligten Bibliotheken wird an der Universitätsbibliothek Leipzig realisiert. Hier laufen auch die Verbindungen zu den Systemen der Projektpartner zusammen.

Der Vorteil, dass alle beteiligten Einrichtungen dasselbe Bibliothekssystem Libero (Version 5.5) verwenden, wird durch den Umstand aufgewogen, dass es sich hierbei um ein sehr geschlossenes System handelt, das nicht annähernd über die benötigten Schnittstellen für eine Anbindung an VuFind (oder jedes andere System) verfügt. Um Bewegungs- und Statusdaten wie die Verfügbarkeit oder Transaktionen im Nutzerkonto in VuFind abzubilden, müssen die Bibliothekssysteme in Echtzeit abgefragt werden. Hierfür wurde eine spezielle Software entwickelt, welche die Zugänge zu den verschiedenen Bibliothekssystemen bündelt und über eine einheitliche Schnittstelle den Discovery Systemen zugänglich macht.

Um auf die verschiedenen Libero-Instanzen zuzugreifen, sind gesicherte Verbindungen zu den Systemen der einzelnen Bibliotheken notwendig. Wegen unterschiedlicher lokaler Gegebenheiten kommen hier verschiedene Technologien zum Einsatz (stunnel, cisco-vpn, ssh). Die Anbindung von Libero erfolgt über eine JDBC-Schnittstelle sowie Webscraping der Web OPAC-Oberfläche. Nach außen kommuniziert die Software als Webservice. Die ermittelten Daten werden im JSON-Format bereitstellt.

Die im SWB geführten Titeldaten werden mit einer täglichen Updateroutine im MARC-Format vom Bibliotheksservice-Zentrum Baden-Württemberg (BSZ) bereitgestellt. Nach der Verarbeitung durch eine im Rahmen des Projekts entwickelte Software, deren Aufgabe das Metadatenmanagement ist, werden die Daten in den Index übernommen. Neben den Titeldaten werden spezifische Bestandsdaten wie Signaturen oder Standorte für jede Einrichtung indiziert und

4 <http://blog.finc.info/kriterienkatalog-zur-ausschreibung-des-aggregierten-index> [28. Mai 2012]

5 http://www.exlibrisgroup.com/de/category/Primo_Central [28. Mai 2012]

6 <http://www.redmine.org/> [28. Mai 2012]

7 <https://katalog.ub.uni-leipzig.de/> [28. Mai 2012]

über das Bibliothekssigel der entsprechenden Instanz zugeordnet. Um diese Daten abzubilden wurde das bibliothekarische Austauschformat MARC21 genutzt und bestimmte Kategorienbereiche für die lokalen Daten vorbehalten.

Ein wesentlicher Vorteil von Discovery Systemen ist, dass neben den in den Lokalsystemen nachgewiesenen Beständen weitere externe Quellen eingebunden werden, die bislang über separate Nachweissysteme erschlossen wurden. Das können zum Beispiel Inhalte des Project Gutenberg, Publikationen aus verschiedenen Hochschulschriftensammlungen oder auch sehr spezielle musikwissenschaftliche Datenbanken sein. Im Projekt *finc* gibt es für die Einbindung externer Daten im Wesentlichen zwei Kriterien: die Verwendung der Daten muss rechtlich zulässig sein und sie müssen in einer nutzbaren Struktur vorliegen.

Um die Datenmengen unterschiedlicher Herkunft und Struktur sauber in den Index zu bringen, zu verwalten, sie anzureichern und korrekt darzustellen, erfolgt die Aufbereitung und zentrale Speicherung der Importdaten in einer klassischen, relationalen Datenbank. Ein

beben und verarbeitet. Über die PC Bridge werden die Resultate zusammengeführt und an das Frontend übergeben. Dabei bietet die PC Bridge Möglichkeiten, die Zusammenführung der Daten nach verschiedenen Parametern zu beeinflussen, sie zum Beispiel für das Ranking unterschiedlich zu gewichten oder unterschiedliche Erschließungsdaten auf gemeinsame Facetten zusammenzuführen (Abb. 2). Um die PC Bridge in realer Anwendung weiter zu optimieren haben die Universitätsbibliothek Leipzig und Ex Libris eine Entwicklungspartnerschaft vereinbart.

Normdaten

Eine weitere Schlüsselfrage im Projekt war der Umgang mit Normdaten. Nachdem die Speicherung der Normdaten in RDF/XML in einem Triplestore sich aus Performancegründen als nicht praxistauglich für den Live-Betrieb herausgestellt hatte, ist nun die folgende Lösung im Einsatz: nach der Umwandlung in MARC/XML werden die Normdaten in einer vereinfachten, tripelähnlichen Datenstruktur in einem Solr-basierten Index gespeichert. Dieser Normdatenindex dient lediglich der schnellen Abfrage von IDs, die auf die Suchanfrage eines Anwenders passen. Der ursprüngliche Suchstring wird nun um die ermittelten IDs erweitert und an die Suchmaschine übergeben.

Auf diese Weise wird nicht nur nach der eingegebenen Zeichenfolge, sondern auch nach der Normdaten-ID gesucht und alle zu dieser ID gehörigen Treffer zurückgegeben (Abb. 3).

Katalogoberfläche

Letztendlich werden die Daten aus allen angeschlossenen Systemen in der Katalogoberfläche, dem Frontend zusammengeführt: Daten aus der Suchmaschine, dem aggregierten Index, den Bibliothekssystemen und externen Diensten werden miteinander verknüpft. VuFind verfügt über die für Discovery Systeme üblichen Grundfunktionalitäten, wie Facettierung, Anzeige ähnlicher Treffer, Kommentarfunktion und Favoritenverwaltung, verschiedene Exportmöglichkeiten oder persistente URLs für Recherchen und Records. Neben diesen Grundfunktionalitäten sind Erweiterungen notwendig, um zum Beispiel die Nutzerdaten (Nutzerkonto) in die Katalogoberfläche zu integrieren und nicht in diesem Punkt wieder auf die Libero Systeme angewiesen zu sein. Dazu kommen lokale Besonderheiten, die aus den Gegebenheiten in den einzelnen Bibliotheken resultieren, wie zum Beispiel die Einbindung von Formularen zur Fernleihbestellung. Im Projekt wird für die beteiligten Bibliotheken eine generische Katalogoberfläche bereitgestellt, die entsprechend den eigenen Bedürfnissen ange-

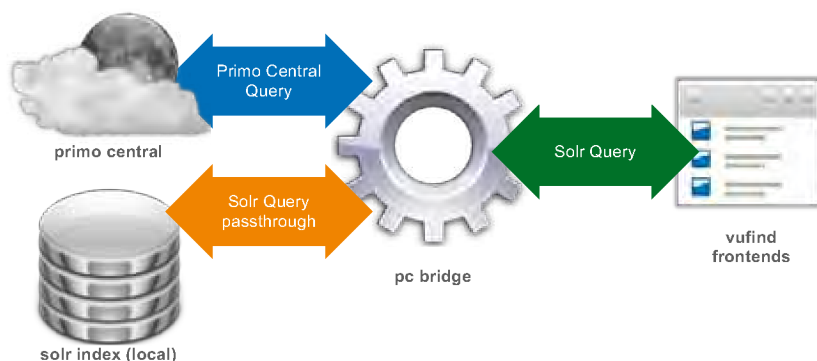


Abbildung 2:
Primo Central
Bridge

effektives Metadatenmanagement entwickelt sich zur zentralen Herausforderung im Projekt (Abb. 1).

Primo Central Bridge

Eine andere Herausforderung ist die Einbeziehung des vom Konsortium lizenzierten aggregierten Index. Ein wesentliches Kriterium bei der Entscheidung für den Index Primo Central der Firma ExLibris war die Option, Suchergebnisse in einer gemeinsamen Ergebnisliste anzuzeigen, ohne dabei die eigenen Daten in einen Mega-Index einzuspielen. Die Software, die dieses umsetzt, ist die sogenannte Primo Central Bridge (PC Bridge), die erstmals 2010 auf der VuFind Conference vorgestellt wurde. Suchanfragen werden durch die PC Bridge an beide angeschlossenen Indizes – dem aus der Metadatenmanagement-Datenbank aufgebauten eigenen Index und dem aggregierten Index Primo Central von ExLibris – weiterge-

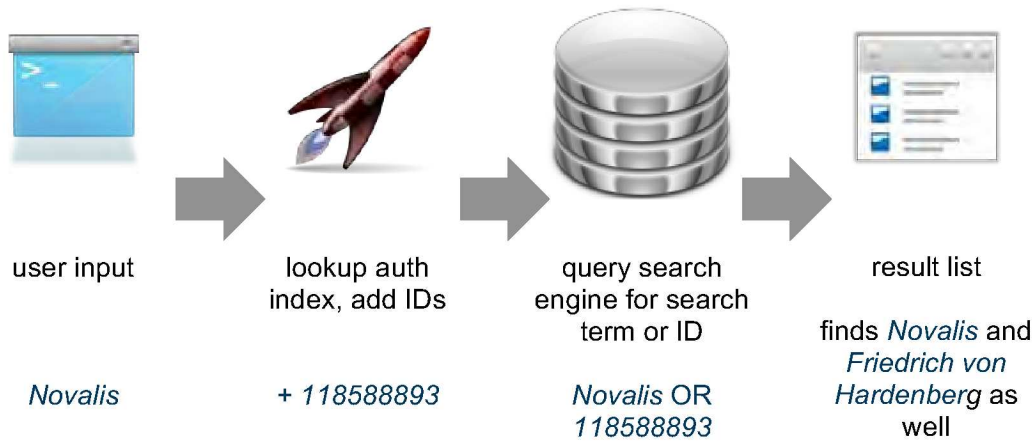


Abbildung 3:
Verwendung von
Normdaten in
VuFind

passt und weiter entwickelt werden kann.

Generell geht es darum, eine Balance zu finden zwischen notwendigen VuFind-Anpassungen und dem Bestreben, die Software konfigurierbar und kompatibel zu halten. Große Abweichungen vom ursprünglichen Quellcode von VuFind sollen vermieden werden, um von weiteren Entwicklungen nahtlos zu profitieren. Alle relevanten Software-Entwicklungen während des Projektes werden unter einer Open Source Lizenz offen zur Verfügung gestellt.

Erste Erfahrungen

Die Freischaltung der ersten Beta-Version eines neuen Kataloges im Projekt *finc* für die Universitätsbibliothek Leipzig, die von der Einführung eines Chat Services und eines Blogs begleitet wird, gibt Gelegenheit, erste Erfahrungen zu sammeln. Insgesamt gab es von Seiten der Nutzerinnen und Nutzer ein positives Feedback. Auch die Tatsache, dass nach kurzer Zeit insgesamt nur noch relativ wenige Rückmeldungen eingingen, darf nach ähnlichen Erfahrungen an anderen Bibliotheken vorsichtig als positives Signal gedeutet werden. Fehlermeldungen, Kritik und Verbesserungsvorschläge kommen vor allem aus dem eigenen Haus und zeugen von aktiver Beteiligung sowie Akzeptanz des neuen Kataloges; die Einführung war mit einer Reihe von Veranstaltungen und Schulungen begleitet worden.

Die mit dem aggregierten Index auf ein Vielfaches gestiegene Datenmenge wurde von Nutzerinnen und Nutzern offenbar nicht als problematisch wahrgenommen. Vereinzelt gab es auch hier Rückmeldungen; eine befürchtete generelle Überforderung konnte nicht festgestellt werden. Zu Nachfragen und Irritationen kam es immer wieder bezüglich der Verfügbarkeit von angezeigten Titeln, da bislang auch rein bibliographische Datenbanken wie zum Beispiel die der Modern Language Association (MLA) einge-

bunden werden und so oft ein Zugriff auf den Volltext oder den entsprechenden Bestand nicht gegeben ist, was von einem Bibliothekskatalog aber offenbar generell erwartet wird.

Ausblick

Nachdem eine stabile Grundfunktionalität des Systems erreicht ist, kann die Gestaltung und Entwicklung weiter vorangetrieben werden. So wird es im Rahmen einer Masterarbeit eine Usability-Studie zum Katalog geben. Während der Projektlaufzeit sollen außerdem die bereits vorhandenen Normdaten angereichert werden, um auch fremdsprachliche Ansatzformen von Personennamen zu berücksichtigen. Auch ist die Entwicklung von Anwendungen für mobile Endgeräte und die Anbindung weiterer Wissensdatenbanken vorgesehen. In den kommenden Monaten aber werden schrittweise zunächst alle beteiligten sächsischen Universitäts- und Hochschulbibliotheken mit der neuen Recherche-Oberfläche ausgestattet werden. In der Universitätsbibliothek Leipzig scheint die Einführung gelungen. **I**



Jens Lazarus

Universitätsbibliothek Leipzig
Bereichsleiter Medienbearbeitung
lazarus@ub.uni-leipzig.de