

Anne Christensen

Katalog 2.0 im Eigenbau: Das beluga-Projekt der Hamburger Bibliotheken

Was macht den Katalog 2.0 aus?

Die Verlagerung von Lehre und Forschung ins Web stellt auch Bibliotheken vor neue Herausforderungen – zumindest dann, wenn sie nach wie vor beanspruchen, eine unverzichtbare Infrastruktureinrichtung für die universitäre Informationsversorgung zu sein. Seit Mitte der 1990er Jahre sind Bibliotheken mit ihrem „Online-Gang“ beschäftigt: Zwar gab es vielerorts bereits elektronische Kataloge, webfähig waren diese jedoch mitnichten. Zudem galt es, die steigende Zahl an webbasierten Ressourcen wie Datenbanken und elektronische Zeitschriften über die ebenfalls noch neuen Web-Präsenzen zugänglich zu machen. Diesen Anforderungen gerecht zu werden bedeutete für viele Bibliotheken einen Kraftakt: Gemessen an den rasant wachsenden Anforderungen kann das Know-how in Bibliotheken nur vergleichsweise langsam weiter entwickelt werden. Zieht man außerdem in Betracht, wie wenig sich die bibliothekarische Arbeit des Sammelns, Erschließens und Bereitstellens von Literatur in den vergangenen Jahrzehnten verändert hat und wie sehr die Einrichtung Bibliothek selbst auf Langfristigkeit angelegt ist, mag man die vielfach beklagte Behäbigkeit bibliothekarischer Webangebote, die mit gänzlich anders aufgestellten Informationsanbietern wie Google, Amazon und Co. konkurrieren, vielleicht gnädiger beurteilen.

Unabhängig davon sind jedoch neue Herausforderungen erwachsen: Bislang stehen die digitalen Angebote und virtuellen Räume von Bibliotheken auf der einen Seite und die von Lehr- und Lernplattformen sowie wissenschaftlichen Fachcommunities auf der anderen Seite weitgehend unverbunden nebeneinander. Die größte Herausforderung für beide Seiten dürfte im Aufbau von serviceorientierten Informationsarchitekturen liegen, auf deren Grundlage Daten und Dienste zwischen den unterschiedlichen Plattformen ausgetauscht werden können: Metadaten und digitale Texte aus der Bibliothek müssen beispielsweise über standardisierte Formate und Protokolle in die einschlägigen Plattformen einfließen können, aus den Plattformen heraus muss ein Rückfluss an User-Generated Content in die bibliothekarischen Angebote möglich sein.

Antwort auf diese neuen Anforderungen will eine neue Generation von Suchportalen geben, die auch als „Next Generation Discovery Tools“ bzw. im deutschsprachigen Raum als „Katalog 2.0“ bezeichnet werden. Der Markt für Katalog 2.0-Produkte besteht aus zumeist Open Source-basierte Eigenentwicklungen von

Bibliotheken einerseits und kommerziellen Lösungen.¹ Die großen wissenschaftlichen Bibliotheken in Deutschland sowie die Bibliotheksverbände haben das Potenzial von Katalog 2.0-Produkten für sich erkannt. An einigen Standorten sind Installationen von kommerziellen Produkten im Echtbetrieb bzw. in Vorbereitung darauf: Die Universitätsbibliotheken in Mannheim und Münster sowie der Kooperativen Bibliotheksverbundes Berlin-Brandenburg (KOBV) setzen auf Primo von ExLibris, an der Universitätsbibliothek Lüneburg wird eine Testinstallation von Touchpoint von OCLC/PICA vorbereitet. Die Eigenentwicklungen an den Universitätsbibliotheken in Bremen, Hamburg, Bochum, Köln, Heidelberg, Karlsruhe und Konstanz sowie an der Verbundzentrale des Gemeinsamen Bibliotheksverbundes in Göttingen liefern zum Teil bereits seit mehreren Jahren wichtige Impulse für die Diskussion um die neuen Informationssysteme. Die unterschiedliche Schwerpunktlegung der einzelnen Eigenentwicklungen erlaubt allerdings nur eine eingeschränkte Vergleichbarkeit der Ergebnisse, zumal Dokumentationen der jeweiligen Quellcodes (noch) nicht verfügbar sind. Die einzelnen Produkte unterscheiden sich teilweise deutlich.

Systemarchitektur

Grundsätzlich ermöglichen die neuen Katalogprodukte eine deutliche Ausweitung der Suchräume. Während der herkömmliche Katalog lediglich selbständige Werke im Besitz einer Bibliothek verzeichnet, wird bei den neuen Lösungen ein Schwerpunkt auf die Integration weiterer Datenbestände, insbesondere Nachweise von Aufsätzen und elektronischen Volltexten, gelegt. Insofern ist das Normieren und Standardisieren von heterogenen Daten eine besondere Herausforderung beim Aufbau der neuen Suchräume, die entweder in einem zentralen lokalen Index zusammengeführt werden oder als parallele Indices im Rahmen einer föderierten Suche befragt werden.

Best Practices:

E-LIB (SuUB Bremen): Portal für Kataloginformationen und elektronische Medien. Eigenentwicklung auf Basis von CiXBase.

Summon (Serial Solutions): Kommerzielle Discovery-Lösung des US-amerikanischen Dienstleisters Serial Solutions für „web scale discovery“, die standardmäßig eine Vielzahl von Aufsatznachweisen enthält, zu denen die lokalen Kataloginformationen einer Bibliothek hinzugeschaltet werden können

Retrievalmethoden

Die neuen Kataloge bieten bislang unterschiedliche Funktionen an, um die thematische Suche zu unterstützen. Dazu gehören das Browsing durch den Gesamtbe-

1 CHRISTENSEN, 2010

stand oder eine Trefferliste oder aber Empfehlungsdienste. Darüber hinaus gibt es Unterstützung bei der Formulierung von Suchanfragen durch AutoCompletion und Rechtschreibprüfungen. Die traditionelle Boole'sche Suche rückt in den Hintergrund.

Best Practices:

Katalog der North Carolina State University: Erster Katalog auf Basis einer Softwarelösung, die nicht ursprünglich für den Gebrauch bibliothekarischen Informationssystemen entwickelt wurde (Endeca). Zahlreiche Drilldown-Möglichkeiten nach formalen, inhaltlichen und Verfügbarkeitskriterien.

finden.nationallizenzen.de: Auf der OpenSource-Lösung vuFind basierende Recherche- und Serviceplattform für national lizenzierte elektronische Ressourcen, bietet u.a. Autocompletion und verzichtet bewusst auf eine erweiterte Suche.

Offenheit

Ein Katalog 2.0 ist von Offenheit in verschiedene Richtungen gekennzeichnet: Die bibliografischen Informationen aus Bibliothekshand können mit zusätzlichen Informationen wie Titelbildern, Inhaltsverzeichnissen und Rezensionen angereichert werden. Darüber hinaus verfügt der Katalog 2.0 über offene, standardisierte Schnittstellen, über die Metadaten und Volltexte in andere Anwendungen übernommen und in verschiedenen Formaten exportiert werden können.

Best Practices:

Kölner UniversitätsGesamtkatalog: Die auf OpenSource-Produkten basierende Eigenentwicklung reichert die nachgewiesenen Metadaten von Zeitschriften u.a. mit Links zu den Inhaltsverzeichnissen an.

OpenLibrary Project: Das Projekt hat das Ziel, für jedes Buch der Welt eine eigene, referenzierbare Website zu erstellen und bietet verschiedene Schnittstellen an, um bibliografische Informationen und Volltexte abzufragen.

Mehrwerte

Katalog 2.0-Produkte verfügen in aller Regel über die Möglichkeit der Personalisierung. Literaturlisten können angelegt und mit anderen geteilt werden. Außerdem werden unterschiedliche Funktionen angeboten, die die User Experience verbessern (Visualisierungen, mobile Dienste).

Best Practices

WorldCat: Gemeinsamer Katalog der Mitgliedsbibliotheken von OCLC. Erreicht durch hohe Bekanntheit insbesondere im angloamerikanischen Raum hohe Nutzungszahlen und eine entsprechend kritische Masse bei kollaborativen Funktionen wie dem Teilen von Literaturlisten.

Libris: An der schwedischen Nationalbibliothek entwickelter Verbundkatalog der schwedischen Bibliotheken, der u.a. mit einer Visualisierung von Verfügbarkeitsinformationen aufwartet.

beluga

Im Rahmen eines gemeinsamen Projektes von sechs wissenschaftlichen Bibliotheken in Hamburg wird unter Federführung der Staats- und Universitätsbibliothek Hamburg am Aufbau einer neuen Rechercheplattform mit dem Titel beluga gearbeitet.² Das Projekt wird vom Hamburger E-Learning Consortium (ELCH) gefördert und hat eine Laufzeit von 36 Monaten (1.11.2007-31.10.2010).

Mit beluga entsteht eine mit Web 2.0-typischen Funktionen ausgestattete Rechercheplattform. Sie vereinfacht die Recherche von 8 Millionen Büchern, Zeitschriften und Aufsätzen in den Hamburger Bibliotheken und versorgt die virtuellen Lernräume und sozialen Netzwerke mit Literaturinformationen und digitalen Texten. Diese Informationen werden in dem sogenannten Hamburg-Index zusammengeführt, der auf einem selbst entwickelten Framework auf Basis der Open-Source-Suchmaschinentechnologie Lucene/SOLR basiert.

Suchmaschinentechnologie

Das zentrale Anwendungsgebiet der Suchmaschinentechnologie in Bibliotheken ist die Indexierung von bibliografischen Metadaten und digitalen Volltexten sowie die Entwicklung von entsprechenden Retrievallösungen wie die Verarbeitung von Suchanfragen und Suchbegriffen oder die Sortierung von Ergebnismengen nach unterschiedlichen Kriterien. Besondere Bedeutung beim Einsatz der Suchmaschinentechnologie hat der Katalog als zentrales Nachweisinstrument für die Bestände einzelner oder mehrerer Bibliotheken. Online Public Access Catalogs oder kurz OPACs gehören seit den 1990er Jahren zum Standardangebot von Bibliotheken. Nach einer kurzen Phase des Angebots von Telnet-OPACs haben die Anbieter von Bibliothekssystemen die so genannten WebOPACs in ihre Produktpaletten aufgenommen, die in der Folge für Benutzerinnen und Benutzer bereitgestellt wurden.

Die OPACs dieser ersten Generation sind jeweils ein Bestandteil von integrierten Bibliothekssystemen: Die Datenquellen werden zwar durchaus auch mit Suchmaschinen indexiert und mit verschiedenen Suchalgorithmen durchsuchbar

2 Neben der SUB Hamburg sind die folgenden Bibliotheken beteiligt: Universitätsbibliothek der Technischen Universität Hamburg-Harburg, Hochschulinformations- und Bibliothekssystem der Hochschule für Angewandte Wissenschaften, Bibliothek der Hochschule für Musik und Theater, Informations- und Medienzentrum der HafenCity Universität, Ärztliche Zentralbibliothek am Universitätskrankenhaus Eppendorf, Fachbibliothek Rechtswissenschaften der Universität Hamburg

gemacht.³ Im Unterschied zu der modernen Suchmaschinenteknologie allerdings sind Datenquellen und Suchmaschine fest miteinander verbunden und Teil einer Gesamtarchitektur, die auch verschiedene andere Aufgaben in der Bibliotheksautomation erfüllt (Ausleihe, Erwerbung, Katalogisierung). Weder die Suchfunktionen noch die Präsentation der Ergebnisse können bei diesen OPACs beeinflusst werden.

Mit dem Aufkommen von kommerziellen und frei verfügbaren Suchmaschinenlösungen ist eine neue, zweite Generation von Katalogen entstanden. Die zu durchsuchenden Datenbestände werden hierfür aus dem integrierten Bibliothekssystem (Integrated Library System, kurz ILS) herausgeholt, in andere, ILS-unabhängige Formate überführt und indexiert.⁴ Es besteht hier keine Bindung mehr zwischen den Datenquellen einerseits und der Suchmaschine andererseits. Dadurch ist es möglich, zusätzlich zu den klassischen Kataloginformationen (Metadaten zu Büchern und Zeitschriften) auch noch weitere Datenquellen in den Index zu integrieren, beispielsweise bibliografische Nachweise von Aufsätzen und elektronischen Ressourcen.⁵ Darüber hinaus können die Retrievalfunktionen und Benutzungsschnittstellen an die individuellen Bedürfnisse von Zielgruppen angepasst und von den Bibliotheken selbst weiter entwickelt werden.

Mit der Entwicklung der Programmbibliothek SOLR für Lucene gibt es seit 2004 eine leistungsfähige Open Source-Indexierungslösung, die speziell auf die Anforderungen der Volltextindexierung ausgerichtet ist und auf Grund der steigenden Anzahl von digitalen Volltexten in den Beständen wissenschaftlicher Bibliotheken eine rasche Verbreitung im internationalen Bibliothekswesen gefunden hat. SOLR ist inzwischen die am häufigsten eingesetzte Technologie für Bibliothekskataloge und teilweise auch in kommerziellen Softwarelösungen für die Indexierung von Daten aus Bibliotheken eingesetzt.⁶ Auch viele Eigenentwicklungen von Bibliotheken basieren auf Lucene/SOLR.⁷

Hamburg-Index: Aufbau und Mehrwerte

Grundlage für den Index von beluga bilden die bibliografischen Daten aus den projektbeteiligten sowie weiteren Hamburger Bibliotheken. Der so genannte „Hamburg-Index“ soll die Bestände aus 14 wissenschaftlichen Bibliotheken der Regionen nachweisen. Dabei handelt es sich um gut 8 Millionen Daten, in der Hauptsache zu Monographien, Zeitschriften und Serien.

3 Diedrichs, 2000

4 Erste Implementierungen von Suchmaschinenteknologie in deutschen Bibliothekskatalogen u.a. an der Universitätsbibliothek Köln (Xapian), der Universitätsbibliothek Karlsruhe (Swish-E), vascoda (FAST)

5 vgl. zum Beispiel E-LIB Bremen, 2010 Bremen: STAATS- UND UNIVERSITÄTSBIBLIOTHEK BREMEN: *Elektronische Bibliothek*. E-LIB. URL <http://elib.suub.uni-bremen.de/> – Überprüfungsdatum 30.03.2010

6 u.a. bei Primo der Firma ExLibris oder bei Touchpoint von OCLC/PICA

7 z.B. HEIDI (UB Heidelberg) oder die Suchkiste (VZG Göttingen)

Die Daten aus heterogenen Bibliothekssystemen werden über einschlägige Schnittstellen eingesammelt und in ein internes Datenformat überführt (Normalisierung). Für die Zusammenführung von bibliografischen Daten unterschiedlicher Herkunft ist eine Festlegung auf ein einheitliches Format notwendig. Dieses Format basiert bei beluga auf Dublin Core (DCMI Metadata Terms). Als internationales Standardformat ist Dublin Core geeignet, die heterogenen Datenbestände für die Indexierung zu vereinheitlichen. Für die Erzeugung von Literaturzitationen übernimmt beluga die Empfehlungen des OpenURL-Frameworks. Die Entscheidung für diesen Weg bei der Normalisierung von Daten fiel auf Grundlage von entsprechenden Vorarbeiten im Rahmen von *vascoda*.⁸

Aktuell werden beluga gestellte Suchanfragen auf verschiedene Weisen vorverarbeitet, u.a. mit Stemming und einer Rechtschreibprüfung. Auf das Angebot einer differenzierten Suchlogik (gefelderte Suche, Phrasensuche, Einsatz von Boole'schen Operatoren) wird derzeit bewusst verzichtet, um den vorrangig studentischen BenutzerInnen von beluga ein schnelles Erlernen der Benutzung zu ermöglichen. Dafür werden in der Ergebnisanzeige mit den Drilldowns für verschiedene formale und inhaltliche Kriterien sowie der Ähnlichkeitssuche Instrumente für die Recherche angeboten, die den eher assoziativen Suchwegen von Studierenden gerecht werden sollen.

Aktuelle Beiträge zu Nutzererwartungen an Bibliothekskataloge⁹ sowie eigene Forschungen im Rahmen der angesprochenen Fokusgruppen und Usability-Studien haben gezeigt, dass BenutzerInnen selbstverständlich von der standardmäßigen Sortierung von Treffermengen nach Relevanz ausgehen, anstatt - wie in konventionellen OPACs üblich - nach Aktualität. Ebenfalls besteht nur eine geringe Bereitschaft zum Durchblättern umfangreicher Ergebnislisten.

Entsprechend dieser Ergebnisse bietet beluga seit der Version 0.6 (Dezember 2008) die Relevanz-Sortierung als Standardeinstellung an.¹⁰ Grundlage für die Relevanzsortierung bei beluga bilden die bibliografischen Daten. In dem für die Version 0.6 entwickelten Algorithmus wurde ein besonderes Gewicht auf die bibliothekarische Verschlagwortung gelegt. Die Aktualität eines Mediums wurde nicht einbezogen, da für entsprechende Einschränkungen die formale Facette "Erscheinungsjahr" in der Ergebnisliste zur Verfügung stand.

Dieser zunächst eher einfach gehaltene Algorithmus ist im Laufe des Förderzeitraumes weiter verfeinert worden. Grundlage dafür bildeten die Analyse und Re-Sortierung von Treffermengen durch Expertinnen und Experten aus den projektbeteiligten Bibliotheken. In der Version 0.9 von beluga (November 2009) wird ein neuer Algorithmus angewendet, in dem dem Erscheinungsjahr ein größeres Gewicht zugesprochen wird. Darüber hinaus werden Titel höher gerankt, die über ein Inhaltsverzeichnis verfügen, da das Vorhandensein eines Inhaltsverzeichnisses in den genannten Studien als zentrales Metadatum für die Beurteilung der Relevanz eines Titels ist. Bevorzugt werden mit dem neuen Algorithmus außerdem

8 Vascoda, 2008

9 ONLINE CATALOGS, 2010

10 CHRISTENSEN, 2008

Werke mit einführendem Charakter, die über einen Kanon aus einschlägigen Begriffen aus dem Titelfeld (z.B. Arbeitsbuch, Textbook, Lehrbuch etc.) identifiziert werden.

Insgesamt haben die bisherigen Bemühungen gezeigt, dass die Erwartungen und Bedürfnisse hinsichtlich einer Relevanz-Sortierung bei den unterschiedlichen Nutzergruppen von beluga (Studierende, Lehrende, technisches Personal und Verwaltungspersonal, Stadtleser) sowie innerhalb von verschiedenen Fachrichtungen ausgesprochen heterogen sind. Dies macht beluga zu einem interessanten Anwendungsfall, um mit neuen Ideen für das Ranking zu experimentieren: Die Hinzuziehung von Ausleih- und Exemplarzahlen sowie eine Bevorzugung von Kernverlagen und Kernzeitschriften sind hierfür interessante Ansätze.¹¹

Entdeckendes Suchen

Mit Rücksicht auf das durch verschiedene Studien belegte, veränderte Rechercheverhalten der „Netzgeneration“¹² bietet die Rechercheplattform neue Wege des Zugangs zu wissenschaftlicher Information: Neben den Möglichkeiten der gezielten Suche nach bekannten Titeln werden unterschiedliche Formen des thematischen Zugangs zu dem in Bibliotheken gespeicherten Wissen angeboten, beispielsweise das facetiierte Browsing durch Treffermengen, das assoziative Entdecken von Literatur nach unterschiedlichen formalen und inhaltlichen Kriterien ermöglicht, sowie ein Browsing-Einstieg und Empfehlungsdienste. Darüber hinaus werden die von BenutzerInnen oft als unzureichend empfundenen bibliothekarischen Metadaten mit weiteren wissenschaftlich relevanten Informationen (Inhaltsverzeichnissen, Rezensionen etc.) angereichert. Dabei erfreuen sich insbesondere die Inhaltsverzeichnisse großer Beliebtheit: Neben den bibliotheksseitig im Rahmen von einschlägigen Scan-Projekten erzeugten Inhaltsverzeichnissen sind hier auch die Verlinkungen zu Google Book Search zu nennen, über die neben dem eigentlichen Inhaltsverzeichnis oftmals auch noch weitere Teile des Volltextes zugänglich sind.

Bibliografische Informationen zu Aufsätzen aus Fachzeitschriften sind aus lizenzrechtlichen und technischen Gründen nicht ohne Weiteres in beluga zu integrieren, obwohl es ein vielfach geäußerter Wunsch von Studierenden und Lehrenden ist, mit einer Rechercheplattform wie beluga Bibliothekskataloge und Fachdatenbanken gleichermaßen zu durchsuchen. Eine solche Suchmöglichkeit ist jedoch zu aufwändig, um sie zusätzlich zu den angestrebten Zielen im Rahmen der Projektlaufzeit zu realisieren. Die Einführung des so genannten Datenbank-Recommendens in Version 0.7 von beluga ist jedoch ein Versuch, um aus beluga als Bibliothekskatalog heraus zumindest auf die Ressourcen in Fachdatenbanken aufmerksam zu machen. In der Trefferliste werden „Aufsatzdatenbanken zu Ihrer Suche“ empfohlen und Direktlinks in diese Datenbanken angeboten. beluga nutzt

¹¹ MAYR, 2010

¹² ROWLANDS et al., 2008

dafür den Fachprofil-Webservice der Elektronische Bibliothek der Staats- und Universitätsbibliothek Bremen: Auf Grundlage einer Analyse von Treffermengen liefert dieser Service die Information darüber, welchen Fachgebieten die Titel mehrheitlich zuzuordnen sind. Für die jeweils dominierenden Fachgebiete werden dann ausgewählte Fachdatenbanken angeboten:

Benutzerinnen und Benutzer von beluga haben allerdings über diese Mehrwerte noch weit hinausgehende Vorschläge dafür gemacht, wie der Katalog 2.0 Wissensknoten abbilden kann: Zum einen wurden intellektuell erstellte Auswahllisten von Literatur zu bestimmten Themen gefordert, wenngleich der akademische Lehrkörper als potenzieller Urheber solcher Auswahllisten diese Verantwortung solcher Aufgaben eher von sich gewiesen hat. Zum anderen wurde gefordert, die Erschließung um eine neue, qualitative Dimension anzureichern, um auf diese Weise beispielsweise Werke mit einführendem oder hoch spezialisierten Charakter von einander zu trennen und für potenzielle Verwendungszwecke auszuzeichnen.

Die Öffnung von eigenen Literaturlisten zur Einsicht für andere hat in den Fokusgruppen, die die Entstehung von beluga begleitet haben, in der Regel für sehr kontroverse Diskussionen gesorgt. Im Projektantrag wurde noch davon ausgegangen, dass für das Teilen von Literaturinformationen innerhalb der Community von Bibliotheksbenutzerinnen und -benutzern eine ausreichende Bereitschaft besteht. Allerdings musste festgestellt werden, dass der Zusammenhang zwischen öffentlichen Listen und daraus generierten Empfehlungsdiensten in der Regel nicht erkannt wird. Darüber hinaus bestehen vielfältige Befürchtungen hinsichtlich eines potenziellen Anonymitätsverlustes oder der unerwünschten Nachnutzung eigener Arbeitsaufwände. Die bei Social-Bookmarking-Diensten zu beobachtenden Synergieeffekte durch gemeinschaftliches Sammeln und Erschließen von Internetseiten lassen sich offenbar derzeit nicht auf einen Bibliothekskatalog übertragen, zumindest nicht einen von der Größenordnung von beluga.

Export und Personalisierung

Die Anbindung von beluga an die Lernmanagement-Systeme (LMS) der Hamburger Hochschule ist eine der Hauptzielrichtungen des durch das E-Learning-Konsortium Hamburg geförderten Projektes. beluga sorgt mit der Bereitstellung von einschlägigen Schnittstellen für die Möglichkeit des Austausches von Daten zwischen Bibliotheken und Informations- und Kommunikations-Infrastrukturen für Lehre und Forschung. Die mit beluga geschaffene Brücke zwischen Bibliotheken und Lernmanagementsystemen, sozialen Netzwerken und Campus-Managementsystemen erlaubt es, bibliografische Informationen und digitale Texte in die genannten Anwendungen zu exportieren und dort weiter zu verwenden.

Um die Lernmanagementsysteme von beluga aus mit Literaturinformationen und digitalen Texten zu versorgen, wurde beluga mit einer Schnittstelle versehen, die diese Informationen in einem standardisierten Austauschformat für E-

Learning-Inhalte ausliefert. Die Literaturlisten werden in Form von Content-Paketen nach dem IMS-Standard übertragen.¹³ Diese Pakete können von allen an den Hamburger Hochschulen eingesetzten LMS (CommSy, Moodle, StudIP und OLAT) entgegengenommen und interpretiert werden. Die Inhalte der Content-Pakete bestehen jeweils aus den Einträgen der Liste im XML-Format, einer XSL-Datei, die den jeweiligen Zitierstil vorschreibt, und ggf. den Volltexten der Literaturangabe. Ergänzt werden die Daten mit den IMS spezifischen Elementen wie z.B. dem IMS-Manifest. Alle Dateien in gezippter Form bilden das IMS-Content-Paket. Auf Seiten der Lernmanagementsysteme sind in der Regel geringfügige Anpassungen nötig, um die Inhalte aus beluga zur Anzeige zu bringen.

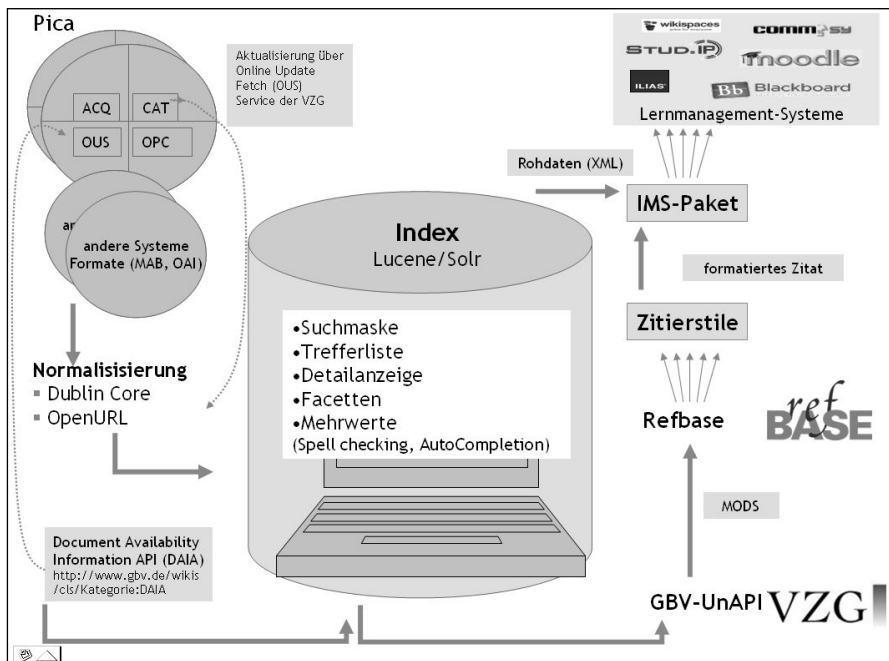


Abb.1: Import und Export von bibliografischen Informationen und Metadaten bei beluga

Um den sehr unterschiedlichen Anforderungen von Lehrenden und Studierenden bezüglich von Zitierstilen und Exportformaten zu begegnen, bedient sich beluga des unAPI-Servers der Göttinger Verbundzentrale des Gemeinsamen Bibliotheksverbundes, um die bibliographischen Daten der antragstellenden Bibliotheken, die für gewöhnlich in dem proprietären Format PICA+ bereitgestellt werden, in das Format MODS zu konvertieren. Aus diesem offenen Format wiederum kann refbase, ein OpenSource-Tool für die Literaturverwaltung, eine große Bandbreite an Zitierstilen und Exportformaten generieren. refbase wurde lokal installiert und

13 IMS, 2010

wird über das im Rahmen von beluga entwickelte und der Entwicklungscommunity zur Verfügung gestellte Perl-Modul Biblio::Rebase mit Daten im MODS-Format gespeist.¹⁴ Für die Ausgabe von bibliografischen Daten stehen damit nun weitere Zitierstile zur Verfügung, darüber hinaus unterstützt beluga jetzt auch Zotero. Mit dieser Infrastruktur werden sich aktuelle und zukünftige Anforderungen von Lehrenden und Studierenden an den Export von Metadaten in unterschiedlichen Zitierstilen und Formaten mit vertretbarem Pflegeaufwand erfüllen lassen.

Die Version 0.8 von beluga, erschienen im Oktober 2009, ermöglicht die Erstellung und Verwaltung von persönlichen Listen sowie den Export dieser Listen in unterschiedlichen Zitierstilen und Formaten. Neben dem neuen persönlichen Bereich „Mein beluga“ gibt es mit den Merklisten und den dazu gehörigen Exportfunktionen jedoch weiterhin die Möglichkeit, beluga quasi anonym zu benutzen und dabei hinsichtlich der Exportfunktionen den gleichen Funktionsumfang zu haben wie in der personalisierten Version. Dort hingegen können die Benutzerinnen und Benutzer ab sofort Literaturangaben („Meine Bibliothek“) sammeln und in Listen anordnen („Meine Listen“). Sowohl für die gesamte Bibliothek als auch einzelne Listen stehen die bisher aus der Merkliste bekannten Exportfunktionen zu Verfügung.

Im persönlichen Bereich von „Mein beluga“ können einzelne Titel elegant per Drag-and-Drop in bestehende Listen gezogen werden. Für die gleichzeitige Bearbeitung mehrerer Titel steht ein Menü zur Verfügung. Für die Nutzung der Personalisierungs-Funktion müssen sich die Benutzerinnen und Benutzer ein Konto bei beluga anlegen. Dafür werden derzeit nur minimale Angaben abgefragt, um einen niedrighschwelligsten Einstieg zu bieten und entsprechende datenschutzrechtliche Vorgaben einzuhalten.

Die Freigabe einzelner Listen zur Einsicht für andere Benutzerinnen und Benutzer ist derzeit noch nicht realisiert. Hier gab es seitens der in den Fokusgruppen befragten Studierenden und Lehrenden unterschiedliche Vorbehalte. Deswegen wurde die Funktion zur Freigabe von Listen in den aktuellen Prototypen nur angedeutet. Eine endgültige Entscheidung über diese Funktionalität wird getroffen werden, wenn eine aussagekräftige Menge an Rückmeldungen zu der Personalisierung im Allgemeinen vorliegen. Neben der persistenten URL für einzelne Titel, die das Abspeichern in Social Bookmarking-Diensten erlaubt, gibt es eine solche persistente URL auch für einzelne Listen.

Vorgehensweise bei der Entwicklung

Bereits bei Antragstellung des Projektes war klar, dass beim Aufbau der Rechercheplattform beluga den Prinzipien der agilen Entwicklung und der partizipativen und dadurch nutzerorientierten Gestaltung gefolgt werden sollte. Grundlage dafür ist die Überzeugung, dass die neue Generation von bibliothekarischen In-

14 Manske, 2008

formationsdiensten nicht nur auf zeitgemäße technologische Standards setzen sollte, sondern bei Konzeption und Umsetzung möglichst nah an den späteren Nutzungsszenarien orientiert sein muss.¹⁵ Die Einbeziehung der künftigen BenutzerInnen von beluga in die Entwicklung und die regelmäßige Überarbeitung der jeweils öffentlich zur Verfügung gestellten Prototypen auf der Grundlage dieser Rückmeldungen sind charakteristisch für beluga. Insofern versteht sich beluga in der Tradition anderer internationaler Katalog 2.0-Projekte, die sich durch intensive Erforschung von User-Bedürfnissen einerseits und Evaluierung der neuen Dienste andererseits auszeichnen.¹⁶

beluga wird von einem multidisziplinären Team aus Bibliothekarinnen, Informatikern und Usability-Expertinnen in der Abteilung IuK-Technik / Digitale Bibliothek der Staats- und Universitätsbibliothek Hamburg entwickelt. Zum Auftakt des Projektes wurde ein Workshop mit deutschen und ausländischen Entwicklerinnen und Entwicklern von ähnlichen Recherche-Lösungen sowie anderen Vertretern der Bibliothek 2.0-Community veranstaltet. Von der dort begonnenen Vernetzung mit anderen Projekten hat beluga in vielfältiger Weise profitiert, beispielsweise durch die Nachnutzung des Fachprofil-Webservice der Bremer E-LIB oder der gemeinsamen Entwicklung des Webservice für Verfügbarkeitsinformationen mit der Göttinger Verbundzentrale des Gemeinsamen Bibliotheksverbundes.

In den ersten Monaten der Projektlaufzeit wurden zudem zwei Fokusgruppen mit Lehrenden der Universität Hamburg durchgeführt, um ein besseres Verständnis davon zu entwickeln, wie diese Zielgruppe nach relevanten Informationen für ihre Lehrveranstaltungen recherchiert und auf welchen Wegen einschlägige Literaturlisten für Studierende zur Verfügung gestellt werden. Die Ergebnisse dieses Workshops flossen in die Entwicklung von Szenarien für den Export von Literaturlisten in Lernmanagement-Systeme ein, die ihrerseits die Richtschnur für die Entwicklung der entsprechenden Funktionalitäten bildeten (vgl. Abschnitt Export und Personalisierung). Für Hintergrundinformationen über das Informationsverhalten der Studierenden wurden unterschiedliche aktuelle ethnografische und kulturanthropologische Studien herangezogen.¹⁷

Der erste Prototyp von beluga wurde im Oktober 2008 öffentlich zugänglich gemacht und sofort mit Hilfe von Usability-Tests von Studierenden auf seine Gebrauchstauglichkeit hin überprüft. Dabei wurde zunächst grundsätzlich bestätigt, was sich in den Fokusgruppen mit den Lehrenden bereits abgezeichnet hatte: Der Bibliothekskatalog wird von beiden Zielgruppen im Wesentlichen als Inventarliste benutzt, um Standorte und Verfügbarkeit von bekannten Titeln zu ermitteln. Der Katalog wird nicht als Entdeckungsinstrument für Literatur zu einzelnen Themen wahrgenommen, was im Gegensatz zum hohen Aufwand für die bibliothekarische Erschließung steht. Allerdings wurde die Idee des nachträglichen Eingrenzens von Treffermengen nach formalen und inhaltlichen Kriterien positiv

15 CHRISTENSEN, 2009

16 u.a. WHITHEAD; TOUB, 2008

17 u.a. GIBBONS, 2007

bewertet, womit sich die Hypothese bestätigt hat, dass die bibliothekarischen Metadaten ihr volles Potenzial noch nicht ausgereizt haben.¹⁸

Auf Grundlage der Usability-Tests, die für die Version 0.7 im Frühjahr 2009 wiederholt wurden, sind in erster Linie Veränderungen in der Benutzungsoberfläche vorgenommen worden, beispielsweise die Umbenennung von einzelnen Facetten oder die prominentere Platzierung der sehr beliebten Links zu Inhaltsverzeichnissen. Entscheidungen von grundsätzlichem Charakter über die Einführung einzelner Funktionalitäten wurden bei beluga in der Regel erst nach Fokusgruppen getroffen. Hierbei bildeten insbesondere die Web 2.0-typischen Funktionalitäten wie das Kommentieren und Bewerten von Literatur, die Öffnung von Literaturlisten oder das Angebot von verhaltensbasierten Empfehlungen einen Schwerpunkt. Anders als erwartet haben diese Funktionalitäten wenig Aussicht auf Akzeptanz. Umso wichtiger sind jedoch die Exportmöglichkeiten in Lernmanagementsysteme und soziale Netzwerke, wo mehr Bereitschaft zur Erstellung solchen User-Generated Contents besteht. Möglicherweise könnte eine Weiterentwicklung des „Katalog 2.0“ aus der Aggregation von solchen nutzergenerierten Metadaten bestehen.

Die Fokusgruppen mit Studierenden haben ebenfalls vielfältige Unsicherheiten in Bezug auf Umfang und Funktionsweise von Katalogen aufgezeigt. Die räumliche und inhaltliche Abdeckung sowie der Anteil von unterschiedlichen Medientypen (Büchern, Aufsätzen, AV-Medien) sind den Benutzerinnen und Benutzern in der Regel unklar. Außerdem fehlt in der Regel eine Rückmeldung über die Korrektheit und Vollständigkeit von benutzten Suchbegriffen. Von einer Sortierung der Treffermenge nach Relevanz wird standardmäßig ausgegangen, obgleich die meisten herkömmlichen Kataloge das Kriterium der Aktualität anwenden. Wenn der Katalog 2.0 intuitiv benutzbar sein will, müssen geeignete Wege gefunden werden, solches Wissen über seinen Umfang und seine Funktionsweise gleichsam beiläufig zu vermitteln, da für formale Wege des Erwerbs entsprechender Informationskompetenz wie die Lektüre von Hilfetexten, den extracurricularen Besuch von Schulungen oder Informationskompetenz-Veranstaltungen die Bereitschaft fehlt. Vielversprechend scheinen hierfür die Möglichkeiten der Visualisierung von Daten zu sein, die bereits in verschiedenen Online-Shopping-Websites im Einsatz sind.

Lessons Learned: Drei Problembereiche

Open Source: Geben ist seliger, aber aufwändiger als nehmen

Mit beluga wurde auf die eigene Entwicklung eines Katalog 2.0 auf Basis von unterschiedlichen Open Source-Technologien gesetzt. Diese Entscheidung wurde vor allem deswegen getroffen, weil bezüglich des Harvestings und der Normali-

18 WIESENMÜLLER, 2010

sierung von heterogenen Metadaten bereits einschlägige Expertise im Haus vorhanden war, die im Rahmen des Aufbaus von unterschiedlichen virtuellen Fachbibliotheken erworben wurde. Zudem sollten die bewilligten Drittmittel nicht in den Erwerb von Lizenzen von kommerziellen Lösungen fließen: Die Eigenentwicklung garantiert eine maximale Gestaltungsfreiheit beim Aufbau der neuen Lösung und verbessert das hausinterne Wissen über neue Such-Technologien und deren Implementierung. Ziel des Projektes ist es, eine an anderen Hochschulstandorten nachnutzbare Lösung zu entwickeln. Mit der Bereitstellung einzelner Komponenten des beluga-Frameworks wurde wie erwähnt bereits begonnen. Im Rahmen der verbleibenden Projektlaufzeit werden sich die Arbeiten allerdings auf die Entwicklung der ausstehenden Funktionalitäten, die Erweiterung des Hamburg-Index und die Ausrichtung des Systems auf den zu erwartenden Lastbetrieb konzentrieren, um zu Ende des Jahres ein lauffähiges System anzubieten, das in der Lage ist, die herkömmlichen Kataloge der beteiligten Bibliotheken zu ersetzen. Erst danach können die Arbeiten an der Dokumentation des Codes abgeschlossen werden, so dass eine Nachnutzung an anderen Standorten möglich ist. Die zeitliche Enge von Drittmittelprojekten wie beluga erlaubt es bedauerlicherweise nicht, der aktiven Beteiligung an der Open-Source-Community in der erforderlichen Weise nachzukommen: Die Nutzung von Open Source-Technologien wird zwar in der Förderlandschaft immer wieder gefordert, aber bei der Bewilligung von Projekten sollte mehr darauf geachtet werden, dass zeitliche und personelle Ressourcen für einen Rückfluss von Programmcode und anderem Know how zur Verfügung stehen, um wirkliche Nachhaltigkeit zu erreichen.

Partizipation: Nicht nur die Benutzerinnen und Benutzer sind gefragt

Rückblickend wäre es sinnvoll gewesen, auf eine stärkere Teilhabe der Kolleginnen und Kollegen in den am Projekt beteiligten Bibliotheken zu bestehen. Der Katalog ist das Herzstück bibliothekarischer Dienstleistungen und braucht auch in der Variante 2.0 das Wissen von Auskunfts- und Katalogabteilungen. Für die bibliotheksfachliche Begleitung von beluga wurde zwar eine Arbeitsgemeinschaft eingerichtet, die sich phasenweise regelmäßig traf, um vor allem über Fragen des Metadatenmanagements und des facettierten Browsings zu diskutieren. Die Anregungen der Benutzerinnen und Benutzer aus den Fokusgruppen und Usability-Studien wurden hier mit großem Interesse aufgenommen. Allerdings fehlte die Zeit für eine ausführlichere Auseinandersetzung mit den verwendeten Methoden und den erzielten Ergebnissen, ebenso wie für einen tieferen Einstieg in Fragen des Retrievals, des Rankings und des Browsings.

Die anfängliche Idee, das Projekt beluga mit einer „beluga-Akademie“ zu begleiten, in der ein erweiterter Kreis von Mitarbeiterinnen und Mitarbeitern aus allen Abteilungen einmal im Monat über Schwerpunktthemen aus der beluga-Entwicklung und dem Hochschulumfeld informiert werden sollte, konnte leider

nicht umgesetzt werden. Insbesondere die offensichtlich veränderte Wahrnehmung des Kataloges durch seine Benutzer und vor allem seine Nicht-Benutzer fordert das Bibliothekspersonal zu einer vertieften Auseinandersetzung heraus. Die Einführung von neuen Suchportalen – unabhängig davon, ob es sich dabei um kommerzielle oder selbst entwickelte Produkte handelt – bedingt, dass sich der Berufsstand über bisherige Erschließungsmodelle und deren mögliche Weiterentwicklung Gedanken macht, um den Bibliothekskatalog nicht zur bloßen Inventarliste verkommen zu lassen. Dies setzt allerdings auch voraus, dass die Implementierung eines neuen Suchportals nicht als alleinige Aufgabe von IT-Abteilungen, sondern als gemeinsames Projekt einer gesamten Bibliothek verstanden wird. Andernfalls wird der bestehende Graben zwischen den klassischen Geschäftsgangs- und Benutzungsbereichen und den IT-Abteilungen unnötigerweise vergrößert.

Das Silo-Problem: Linked Open Data als Lösung

Wie viele andere der neuen Suchportale auch ist beluga darauf angewiesen, Metadaten und Verfügbarkeitsinformationen aus ihren ursprünglichen Umgebungen herauszuholen: Bei einem Katalog 2.0 handelt es sich oftmals um eine zusätzliche Recherche- und Service-Schicht für Daten, die traditionellerweise in einem integrierten Bibliothekssystem (z.B. LBS3 von OCLC/PICA, Aleph von ExLibris) erzeugt und verwaltet werden. Die Weiterverwendung dieser Daten in einer anderen Umgebung ist im Grunde nicht vorgesehen: Zwar verfügen die Systeme über bibliotheksspezifische Schnittstellen und Austauschprotokolle, den Anforderungen des Katalog 2.0 genügen diese jedoch oft nur unzureichend, da dessen Systemarchitektur in der Regel auf anderen, im Web weiter verbreiteten Standards aufbaut. Ein weiteres Problem sind die vielfältigen Verknüpfungen zwischen verschiedenen Datensätzen sowie deren hierarchischer Aufbau (z.B. im Falle von mehrbändigen Werken). Diese Verknüpfungen sind üblicherweise insofern proprietär, als sie auf internen Identifikatoren beruhen. Eine Nachnutzung dieser Verknüpfungen im Kontext eines Katalog 2.0 ist deswegen mitunter schwierig. Nicht nur aus diesem Grund sollten sich alle Verantwortlichen für Katalog 2.0-Projekte den Forderungen nach der Bereitstellung von bibliografischen Informationen als Linked Open Data anschließen.¹⁹ Insbesondere die von der Deutschen Bibliothek gepflegten Normdaten können für eine automatische Suchterweiterung herangezogen werden und dem Wunsch vieler Benutzerinnen und Benutzer nach einer Rückmeldung über die Angemessenheit der verwendeten Suchbegriffe Rechnung tragen. Voraussetzung dafür ist das Angebot von Webservices, über die Norm- und andere Daten strukturiert abgefragt werden können.

Mit der Forderung nach Linked Open Data geht auch einher, dass Anbieter von bibliografischen Metadaten ihre Geschäftsmodelle überdenken müssen, auf deren Grundlage sie Bibliotheken ihre Daten anbieten. Einzelne Bibliotheken wie

19 POHL, 2010

die des CERN²⁰ oder des Hochschulbibliotheksentrums in Köln²¹ haben bereits ihre Daten kostenfrei unter einschlägigen Lizenzen im Netz zur Verfügung gestellt. Der Gemeinsame Bibliotheksverbund hat in seinem aktuellen Strategiepapier definiert, dass von seinen Mitgliedern erzeugte Metadaten frei zur Verfügung gestellt werden sollen.²² Mit Blick auf die weitere Entwicklung von bibliothekarischen Rechercheinstrumenten und der Möglichkeit der Verlinkung von bibliografischen Daten mit anderen Inhalten im Web handelt es sich dabei um zukunftsweisende Schritte: Der Bibliothekskatalog ist in Zukunft nur eine von vielen Umgebungen, in der bibliografische Informationen aus Bibliotheken verwendet werden. Die Nachnutzbarkeit von Metadaten und Normdaten als bibliothekarischen Kernprodukten in unterschiedlichen Kontexten muss durch neue Datenmodelle wie Linked Open Data sowie Lizenzen wie Creative Commons gefördert werden.

Literaturverzeichnis

- (CERN, 2010) CERN SCIENTIFIC INFORMATION SERVICE: *CERN Library bookdata*. URL <http://library.web.cern.ch/library/Library/bookdata.html> – Überprüfungsdatum 30.03.2010
- (CHRISTENSEN, 2008) CHRISTENSEN, Anne: *beluga 0.6: Die Weihnachts-Edition mit geänderter Backmischung*. URL <http://beluga-blog.sub.uni-hamburg.de/blog/2008/12/17/beluga-06-die-weihnachts-edition-mit-geaenderter-backmischung/>. – Aktualisierungsdatum: 17.12.2008 – Überprüfungsdatum 30.03.2010
- (CHRISTENSEN, 2009) CHRISTENSEN, Anne: *Partizipative Entwicklung von Diensten in der Bibliothek 2.0: Methoden und Ergebnisse aus Katalog 2.0-Projekten*. In: *Bibliotheksdienst* 43 (2009), Nr. 5, S. 527–537. URL http://www.zlb.de/aktivitaeten/bd_neu/heftinhalte2009/Erschliessung010509BD.pdf – Überprüfungsdatum 2010-03-30
- (CHRISTENSEN, 2010) CHRISTENSEN, Anne: Best Practices. URL <http://katalog2null.wordpress.com/2010/01/05/best-practices/>. – Aktualisierungsdatum: 05.01.2010 – Überprüfungsdatum 30.03.2010
- (DIEDRICHS, 2000) DIEDRICHS, Reiner: PSI : Pica Search and Index als Basis für bibliothekarische Informationssysteme. URL <http://www.gbv.de/vgm/info/biblio/01VZG/06Publikationen/2000/pdf/summit00-diedrichs.pdf>. – Aktualisierungsdatum: 15.11.2000 – Überprüfungsdatum 30.03.2010
- (GBV, 2010) GBV: *Strategien für die Informationsversorgung der Zukunft im Spannungsfeld sich verändernder Informationsstrukturen : Strategische Planungen des GBV 2011-2015*. URL http://www.gbv.de/vgm/info/biblio/02GBV/PDF/PDF_4128.pdf. – Aktualisierungsdatum: 1.10.2010 – Überprüfungsdatum 30.03.2010
- (GIBBONS, 2007) GIBBONS, Susan: *Studying students : The Undergraduate Research Project at the University of Rochester*. FOSTER, Nancy Fried (Hrsg.). Chicago : Association of College and Research Libraries, 2007. – ISBN 0838984371

20 CERN, 2010

21 HBZ, 2010

22 GBV, 2010

- (HBZ, 2010) HBZ: *Linked Open Data*. URL http://www.hbz-nrw.de/projekte/linked_open_data/. – Aktualisierungsdatum: 15.03.2010 – Überprüfungsdatum 30.03.2010
- (IMS, 2010) IMS Global Learning Consortium: IMS GLC: Content Packaging Specification. URL <http://www.imsglobal.org/content/packaging/>. – Aktualisierungsdatum: 04.01.2010 – Überprüfungsdatum 30.03.2010
- (Manske, 2008) Manske, Henning: Biblio :: rebase. URL <http://www.refbase.net/index.php/Biblio::Rebase>. – Aktualisierungsdatum: 2008-11-30 – Überprüfungsdatum 30.03.2010
- (MAYR, 2010) MAYR, Philipp: *Bradfordizing mit Katalogdaten : Alternative Sicht auf Suchergebnisse und Publikationsquellen durch Re-Ranking*. In: *BuB* 62 (2010), Nr. 1, S. 61–63
- (ONLINE CATALOGS, 2010) ONLINE CATALOGS: WHAT USERS AND LIBRARIANS WANT: *An OCLC Report*. URL <http://www.oclc.org/us/en/reports/onlinecatalogs/fullreport.pdf> – Überprüfungsdatum 30.03.2010
- (POHL, 2010) POHL, Adrian: *Dimensionen von Open Bibliographic Data*. URL <http://www.uebertext.org/2010/03/dimensionen-von-open-bibliographic-data.html>. – Aktualisierungsdatum: 25.03.2010 – Überprüfungsdatum 30.03.2010
- (ROWLANDS ET AL., 2008) ROWLANDS, I. Nicholas D. Williams P. Huntington P. Fieldhouse M. Gunter B. et al: The Google generation: The information behaviour of the researcher of the future. In: *Aslib Proceedings: New Information Perspectives* 60 (2008), Nr. 4, S. 290–310
- (vascoda, 2010) vascoda: vascoda Application Profile Version 2.1 : Zur Standardisierung von Metadatenlieferungen an vascoda. URL http://edok01.tib.uni-hannover.de/edoks/e01vascoda/Arbeitspapiere/vascoda-AP_2.1_Dez-2008.pdf. – Aktualisierungsdatum: Dezember 2008 – Überprüfungsdatum 30.03.2010
- (Whithead; Toub, 2008) WHITHEAD, Martha; TOUB, Steve: *User-Generated Content and Social Discovery in the Academic Library Catalogue: Findings From User Research*. URL <http://www.slideshare.net/stoub/usergenerated-content-and-social-discovery-in-the-academic-library-catalogue-findings-from-user-research-presentation>. – Aktualisierungsdatum: 2.10.2008 – Überprüfungsdatum 30.03.2010
- (WIESENMÜLLER, 2010) WIESENMÜLLER, Heidrun: *Metadaten härter arbeiten lassen und besser präsentieren: Sacherschließung und Normdaten in Online-Katalogen*. In: *BuB* 62 (2010), Nr. 1, S. 48–54