

## Green Open Access im Bibliothekskatalog - Chancen & Risiken

Martin Blenkle, Rachel Ellis, Elmar Haake, Manfred Nölte

**Staats- und Universitätsbibliothek Bremen**

Bibliotheken bewerben seit etwa 15 Jahren intensiv Open Access Modelle als neue offene Publikationsformen im Wissenschaftsbereich. Dennoch wird das Thema eines nutzerfreundlichen Nachweises solcher Open Access Dokumente in Bibliothekskatalogen immer noch kaum beachtet. Um die Akzeptanz grüner Publikationsmodelle nachhaltig zu fördern, ist es jedoch sinnvoll, Nutzer\*innen die wertvollen Nachweise der frei verfügbaren wissenschaftlichen Dokumente auch im Gesamtkontext bibliothekarischer Kataloge bei der Suche anzubieten. Sollen die Förderung von Open Access Texten und die Gleichstellung mit lizenziertem Material aus klassischen Publikationsformen wirklich ernsthaft betrieben werden, ist die Integration in die Standardnachweissysteme der Bibliotheken unausweichlich, um die Sichtbarkeit dieser Medien für den Wissenschaftsbetrieb zu erhöhen. Dagegen besteht bisher immer noch vorherrschend das Konzept des universitären Dokumentenservers als Inselangebot mit eigener Suchoberfläche, das erfahrungsgemäß von Nutzer\*innen aber nur selten direkt angesprochen wird.

Die Recherche über ein breites Gesamtangebot an grünen Open Access Medien ist bisher nur über gängige Internetsuchmaschinen sowie über die Spezialsuchmaschine BASE der UB Bielefeld möglich. Eine integrierte bibliothekarische Gesamtsicht auf lokal vorhandene Printmedien, lokal lizenzierte E-Medien und freies Open Access Material ist bundesweit in den Bibliotheken bisher kaum verfügbar. Zwar stehen die Metadaten von weltweit derzeit etwa 4300 Dokumentenservern<sup>1</sup> seit etwa 15 Jahren via OAI-PMH Schnittstellen prinzipiell frei zur Nutzung bereit, doch scheint es viele Vorbehalte gegen die Integration solcher Nachweise in ein gemeinsames Angebot mit bibliothekarisch kontrollierten Metadaten zu geben.

So ergibt sich die Situation, dass viele Bibliotheken erhebliche Aufwendungen in die Förderung von Open Access Publikationsformen investieren und damit wertvolle Überzeugungsarbeit an den Universitäten leisten, gleichzeitig aber die Auswirkungen dieser Maßnahmen in den bibliothekarischen Nachweisangeboten kaum sichtbar sind.

An der Staats- und Universitätsbibliothek (SuUB) Bremen ist der Nachweis frei verfügbarer wissenschaftlicher Texte bereits seit 2004 ein wichtiger Teil der Marketingstrategie von Open Access Publikationsmodellen. Das Angebot an Open Access Dokumenten im Rahmen der zentralen Bibliotheksuchmaschine erfolgt über das bereits seit 2004 an der SuUB Bremen verfügbare eigene Discoverysystem E-LIB.

---

<sup>1</sup> 4.285 Dokumentenserver – ca. 90 Mio. Nachweise Quelle BASE (Stand:21.04.2016)  
[https://www.base-search.net/about/de/about\\_sources\\_date\\_dn.php?menu=2](https://www.base-search.net/about/de/about_sources_date_dn.php?menu=2) (Aufruf 21.04.16)

## **Die Elektronische Bibliothek Bremen (E-LIB) – ein Entwicklungssystem für neue Katalogdienste für wissenschaftliche Bibliotheken**

Die E-LIB Bremen weist im Sinne moderner Discovery-Systeme E-Medien und Printmedien in einem lokalen Index nach und präsentiert die Titel gemeinsam in einer integrierten Trefferliste. Schon seit 2004 stehen auch Nachweise von Volltexten grüner Open Access Archive zur Verfügung. Die Open Access Daten werden dabei gemeinsam mit den bekannten Artikelangeboten klassischer Verlage völlig gleichberechtigt im Katalog angezeigt. Seit 2010 ersetzt das System den bisherigen klassischen Bibliotheks-OPAC als Zugangssystem komplett und wickelt derzeit etwa 13.000 bis 18.000 Suchanfragen täglich ab.<sup>2,3</sup> Die E-LIB Bremen versteht sich dabei gleichzeitig als Produktivsystem und Entwicklungsumgebung für neue Services zur Unterstützung der Literatursuche von Nutzer\*innen.

Im Gesamtindex der Bremer E-LIB finden sich derzeit ca. 60 Mio. Nachweise – davon mehr als 95% elektronische Volltexte. Die eingesetzte Katalogtechnik basiert auf einem lokal gepflegten Gesamtindex von heterogenen, aggregierten Metadaten. Der Index der Suchmaschine wird durch die Integration weiterer Metadatenansammlungen im Sinne eines Data-Warehouse kontinuierlich erweitert. Die Auswahl der integrierten Metadatenbestände erfolgt nach inhaltlichen Kriterien und dient vordringlich der gezielten Informationsversorgung für den Standort der Universität und Hochschulen im Lande Bremen. Im Gegensatz zu der in Discoverysystemen meist eingesetzten Indexierungs- und Retrievalsoftware Apache Lucene nutzt die E-LIB Bremen das System CiXbase<sup>2</sup> zur internen Organisation der Metadaten, ein Indexierungs- und Recherchesystem unter Nutzung computerlinguistischer Verfahren.

---

<sup>2</sup> Blenkle, Martin; Ellis, Rachel; Haake, Elmar: Next-generation library catalogues: review of E-LIB Bremen. In: *Serials: The Journal for the Serials Community* 22 (2009), H. 2, S. 178-181.

<http://uksg.metapress.com/content/040q4504p2565519?id=040q4504p2565519> (Aufruf 21.04.16).  
Die Bibliothekssuchmaschine E-LIB ist in die Webpräsenz der SuUB Bremen vollständig integriert und verfügbar via <http://www.suub.uni-bremen.de/> (Aufruf 21.04.16).

<sup>3</sup> Blenkle, Martin; Ellis, Rachel; Haake, Elmar: Only the first results count: user-feedback-modified relevance ranking in E-LIB Bremen. In: *Insights* 28 (2015), H. 2, S. 75–80. <http://doi.org/10.1629/22178> (Aufruf 21.04.16)

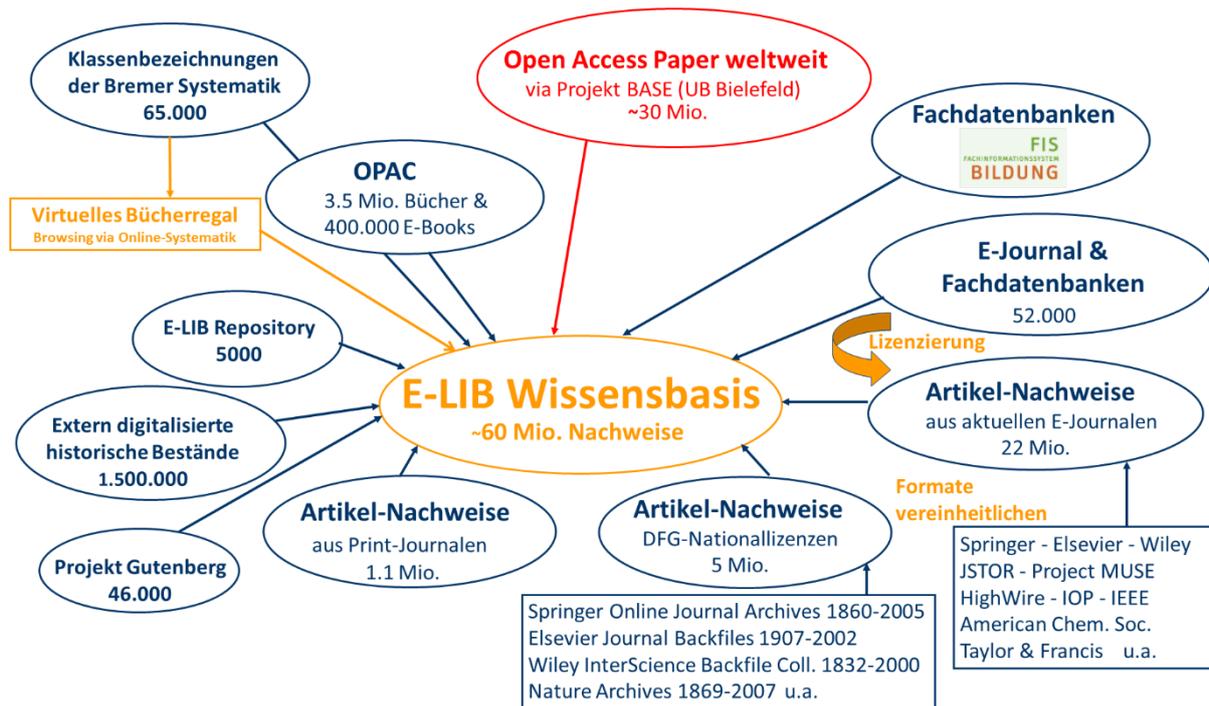


Abbildung 1: Wissensbasis der E-LIB Suchmaschine als Sammlung heterogener Metadatenpakete u.a. Open Access Nachweise

### Integration von Metadaten von weltweit verfügbaren Open Access Medien

Seit der Bildung der Open Archives Initiative (OAI) im Jahr 2000 und dem zur Weitergabe von Archivmetadaten festgelegten, standardisierten Verfahren (OAI Protocol for Metadata Harvesting -OAI-PMH) ist es für Serviceanbieter möglich, Metadaten weltweiter Archive zu sammeln (harvesten) und Nutzer\*innen zur Recherche anzubieten.

An der Staats- und Universitätsbibliothek wurde bereits im Jahr 2004 damit begonnen, den Suchindex der E-LIB um die Nachweise einzelner, für den lokalen Standort gezielt ausgewählter Archive selbst teilweise zu harvesten. Dazu wurden die von etwa 80 Repositorien<sup>4</sup> gesammelten Metadaten in den Suchindex der E-LIB integriert (ca. 3.5 Mio.) und neben dem lizenzierten Material größerer Verlagsanbieter bei der Recherche mitangeboten. Im Rahmen der experimentellen Nutzung ausgewählter Archive konnten viele nützliche Erfahrungen im Metadatenmanagement – sowohl beim Sammeln und aggregieren als auch beim Anbieten von Metadaten freier Dokumente innerhalb der Katalogsicht – gesammelt werden.

<sup>4</sup> Für das Harvesten ausgewählt wurden damals regionale, nationale und internationale Repositorien wie z.B. Alfred-Wegener-Institut Bremerhaven, größere deutsche und internationale Universitätsarchive, CERN, ETH Zürich, PubMed- und BioMedCentral, ArXiv.org, CiteSeer, Cogprints, PhysNet u.a.

Obwohl es bereits früh erkennbar war, dass das Angebot an zusätzlichen Online Medien von den Nutzer\*innen gut angenommen wurde, zeigte die Pilotphase, dass der investierte Aufwand an Personalressourcen nicht unerheblich war. Eine Ausweitung des Angebotes durch Vergrößerung der Anzahl an geharvesteten Archiven kam daher aufgrund der verfügbaren Personalkapazität zunächst nicht in Frage.

Eine deutliche Vergrößerung der Nachweisbasis an OAI Metadaten im Angebot der E-LIB gelang erst durch Nachnutzung der Metadatenammlung der wissenschaftlichen Suchmaschine BASE<sup>5</sup>, einem Projekt der UB Bielefeld. BASE verfolgt das Ziel, weltweit Metadaten aller wissenschaftlichen Repositorien, die Daten im Rahmen des Standards der Open Archives Initiative anbieten, zu sammeln, zu aggregieren und in einer Suchmaschine zur Recherche anzubieten.

Im Rahmen einer Kooperation steht seit 2012 der gesamte Metadatenbestand des Projekts BASE für die E-LIB zur Verfügung – insgesamt ca. 90 Mio. Metadaten. In der Bremer Suchmaschine wird derzeit eine Teilmenge von etwa 25 Mio. Nachweisen<sup>6</sup> aus dem Gesamtbestand von BASE im integrierten Gesamtangebot von lokalen Printtiteln und lizenziertem Online-Material größerer Verlage angeboten (siehe beispielhaft Abbildung 2).

---

<sup>5</sup> "BASE (Bielefeld Academic Search Engine) ist eine der weltweit größten Suchmaschinen speziell für frei im Sinne des Open Access zugängliche wissenschaftliche Dokumente im Internet." Derzeit umfasst die BASE Sammlung ca. 90 Mio. Nachweise von 4300 Repositorien – davon sind geschätzt 60% frei verfügbar (Open Access) (Stand April 2016). <https://www.base-search.net/about/de/index.php> (Aufruf 21.04.16)

<sup>6</sup> Die Gesamtzahl an importierten BASE-Daten in der E-LIB Bremen schwankt kontinuierlich mit jeder Aktualisierung aufgrund der Filterung des Datenmaterials (siehe unten).

**b** Staats- und Universitätsbibliothek Bremen

Ausweisnummer  
.....>  
Nutzerdaten vergessen?  
Jetzt Konto anlegen und registrieren  
(hier geht's zur Medienverlängerung)

Über uns Service & Beratung Kataloge & Sammlungen Fachinformationen Standorte Kontakt Literatur verwalten & publizieren

Home a a

Suche Alle Medien César Manrique

46 Ergebnisse:

## César Manrique

Hilfe Ergebnis drucken Suche als RSS

Ergebnisse 1-15 Sortierung: Relevanz

**ONLINE**  
UR BASE

**Art and nature in Lanzarote : the influence of César Manrique...**  
Pezzi, Maria Giulia , 2013  
Link: <http://resolver.obvsg.at/urn:nbn:at:at-ubg:1-52602>

**ONLINE**  
UR BASE

**Luces y sombras de la vanguardia artistica canaria: diálogo...**  
Nuez Santana, José Luis de la , 2015  
Link: <http://mdc.ulpgc.es/cdm/ref/collection/aea/id/2313>

**ONLINE**  
UR BASE

**César Manrique : obra espacial:**  
Calabuig, Tino; Ojesto, Pedro; Chacón, Joaquín , 1995  
Link: <http://mdc.ulpgc.es/cdm/ref/collection/mdcm/id/179>

**ONLINE**  
UR BASE

**César Manrique : the game is the message:**  
Castro, Fernando , 1992  
Link: <http://mdc.ulpgc.es/cdm/ref/collection/atlantica/id/39>

**ONLINE**  
UR BASE

**César Manrique : el juego es el mensaje:**  
Castro, Fernando , 1992  
Link: <http://mdc.ulpgc.es/cdm/ref/collection/atlantica/id/23>

**Ergebnis eingrenzen**

**Medientypen**  
Bücher(2)  
OpenAccess-Volltexte (44)

**Sprachen**  
deutsch (1)  
englisch (4)  
mehr...>>

**Standort der Titel**  
Uni Bremen (2)

**Autor(en):**

**Jahre eingrenzen**

1970 bis 2015

4 3 20 6 12

1970-1979 1979-1988 1988-1997 1997-2006 2006-2015

Abbildung 2: Literatur über César Manrique in der E-LIB Bremen: 2 Monographien und 44 Open Access Texte aus dem Metadatenpaket BASE – freie Open Access Texte werden besonders gekennzeichnet und können über eine Facettierungsoption gezielt angewählt werden.

## Open Access Medien im Bibliothekskatalog / Nutzungsaspekte

Durch die mehrjährigen Erfahrungen beim Anbieten freier Dokumente im Bibliothekskatalog ergeben sich verschiedene neue Aspekte in Fragen der Bereitstellung, aber auch Probleme bei der Nutzung dieses Medientyps, die zu organisatorischen Veränderungen und zur Entwicklung neuer Supportdienste führen können.

## Nutzungshäufigkeit von Open Access Material in der Suchmaschine E-LIB

Eine Analyse der Nutzungshäufigkeit der importierten Metadaten im Vergleich zu anderen Angeboten des gemeinsamen Suchraumes der E-LIB Bremen ist nicht ganz einfach, da alle vergleichbaren Datentypen in unterschiedlicher Anzahl vorgehalten werden. Erschwerend kommt hinzu, dass die Metadaten eines jeweiligen Typs je nach Importkonditionen monatlich in schwankender Anzahl indexiert werden<sup>7</sup>.

Unter Berücksichtigung dieser Rahmenbedingungen erhält man einen ersten Hinweis auf die Nutzung der freien Inhalte durch den Vergleich der Gesamtnutzung aller angebotenen, insbesondere der lizenzierten E-Medien mit der Nachfrage nach den freien BASE/OAI-Angeboten (siehe Abbildung 3).

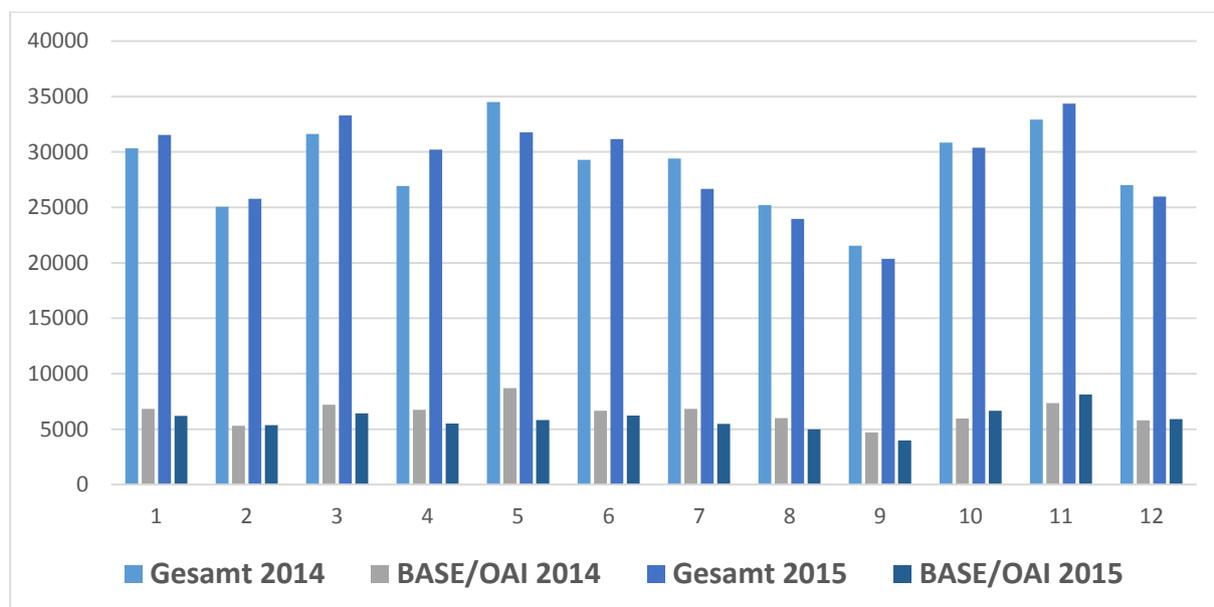


Abbildung 3: E-LIB Bremen: Informationen zur Nutzung freier Open Access Medien (Trend): Anteil an monatlichen Zugriffen auf freie Texte aus BASE/OAI-Metadaten im Vergleich zum Gesamtzugriff auf alle E-Ressourcen 2014/2015

Etwa 20% der E-Mediennutzung entfiel im Betrachtungszeitraum der letzten zwei Kalenderjahre auf die neuen kostenfreien Angebote – bei einer durchschnittlichen Nutzung von etwa 1000 täglichen Zugriffen auf alle E-Medien, lag der Anteil der Zugriffe auf BASE/OAI Daten damit bei etwa 200 am Tag.

<sup>7</sup> Weiterhin sind nicht alle Indextypen der E-LIB Bremen - historisch bedingt - sauber differenzierbar: Die Metadaten des lokalen Bremer Bibliothekssystems (PICA) enthalten zwar überwiegend Printmaterial, aber auch zahlreiche E-Medien wie etwa E-Books und Metadaten von unselbständiger Literatur. Die Nutzung dieser E-Ressourcen bleibt bei dieser Betrachtung unberücksichtigt.

## **Probleme durch Metadaten mit falscher Open Access Zuordnung**

Die Qualität der von vielen Anbietern bereitgestellten Metadaten bereitet den Anbietern von Suchdiensten in bestimmten Bereichen Probleme. Dies betrifft die Nachweise sowohl deutscher als auch weltweiter Open Access Archive - bezieht sich aber weniger auf klassische Qualitätsdefizite der formalen Beschreibung der Medien, sondern vielmehr auf den korrekten Status ihrer Verfügbarkeit.

Ein signifikanter Anteil der im Kontext der Open Archive Initiative von Repositorien angebotenen Metadaten beschreiben Medien, die nicht allgemein frei verfügbar sind und zudem oft auch keine Informationen über diesen Status ausweisen. Viele Repositorien dienen offensichtlich zugleich auch als lokale bibliografische Sammlungen, etwa zum Aufbau einer Universitätsbibliografie und weisen somit durchaus auch lizenzpflichtiges Material nach. Aus diesem Grunde finden sich in diesen Archiven Metadaten lizenzpflichtiger Verlagsangebote, die nicht frei zugänglich und mit der Grundphilosophie eines Repositoriums im Sinne der Open Access Initiative kaum in Einklang zu bringen sind. In vielen Fällen werden diese Daten dann über die OAI-PMH Schnittstellen irrtümlich als "frei verfügbar" definiert an die Datenharvester ausgegeben. Suchdiensteanbieter wie BASE oder E-LIB Bremen können diese Nachweise daher nicht sicher erkennen und gesondert anzeigen. Für Nutzer\*innen von Suchmaschinen- und Discoverykatalogen sind diese *Pseudo-Open Access* Nachweise in der Regel nahezu wertlos.

Es gibt derzeit Ansätze auf Seiten der Dienstanbieter, solche fälschlich als frei verfügbar erhaltenen Metadaten durch intellektuelle Prüfung einzelner Repositorien in verschiedene Open Access Kategorien einzustufen und dann entsprechend zu behandeln. BASE kategorisiert Metadatenlieferanten nach ihrem prozentualen Anteil an frei verfügbarem Material und kennzeichnet dies in den Daten mit einem Attribut. An der Staats- und Universitätsbibliothek wird beim Import der BASE Daten dieses OA-Attribut ausgewertet. Archive mit einem geringen Anteil an frei verfügbaren Angeboten werden nicht in den E-LIB-Gesamtindex importiert und auf diese Weise ausgefiltert. Die Filterung ist nur auf Ebene der Archive möglich. Viele Archive mit einem hohen Anteil von lizenzpflichtigen Medien weisen jedoch auch viele freie Dokumente auf. Solange jedoch eine valide Kennzeichnung der einzelnen Medien nicht verfügbar ist, gehen viele interessante freie Nachweise durch ein zu restriktives Ausfiltern des gesamten Archivs verloren.

Viel sinnvoller wäre eine eindeutige Beschreibung der Medienverfügbarkeit in den von den Datenanbietern bzw. den Repositorien über die OAI-Schnittstellen ausgelieferten Metadaten.

Die Systeme sollten entweder nur Metadaten freier Dokumente über ihre Schnittstellen ausliefern, wie dies ursprünglich im Rahmen der Open Archive Initiative für offene Archive gedacht war. Falls dies nicht umsetzbar ist, sollten die Datenanbieter wenigstens ein entsprechendes Subset von Nachweisen nur freier Medien anbieten, wie dies im DINI Zertifikat 2013 als Mindestanforderung formuliert wird<sup>8</sup>. Zwar ist es sehr sinnvoll, die deutliche Kennzeichnung der Medienverfügbarkeit in den deutschen Standards zu verankern, eine wirkliche Verbesserung der Situation ist aber vorerst nicht in Sicht, da die weltweiten Archive sich kaum an derartigen nationalen Standards orientieren.

Um den Nutzer\*innen möglichst wenig Nachweise mit unklaren Zugriffskonditionen anzubieten, werden die Metadaten vor der Integration in den Suchraum der E-LIB umfangreich analysiert und gefiltert (siehe unten).

### **Usability Probleme im Bereich der Medienbeschreibung / Repositorien Frontdoor Seite**

Im Verlauf der Bereitstellung der Daten im Bibliothekskatalog machen sich weitere Nutzungshemmnisse beim Support bemerkbar, die durch ungünstige Gestaltung und fehlende Nutzerfreundlichkeit vieler Repositorien verursacht werden. Nach den in Bremen gemachten Erfahrungen im Zusammenhang mit den Dokumentenservern unterschiedlichster Anbieter werden Aspekte der Usability und der User Experience beim Design bibliothekarischer Repositorien noch zu wenig beachtet. Die oft ungünstig gestalteten Dokumenten-Frontdoor-Seiten<sup>9</sup> zahlreicher Repositorien führen immer wieder zu Nachfragen.

---

<sup>8</sup> Die AG Elektronisches Publizieren der deutschen Initiative für Netzwerkinformation erarbeitet mit dem DINI-Zertifikat, das inzwischen in der Version 2013 vorliegt, nationale Standards für den Betrieb wissenschaftlicher Dokumentenserver. <http://nbn-resolving.de/urn:nbn:de:kobv:11-100217162> (Aufruf 21.04.16). Aus dem **DINI Zertifikat 2013** S. 37: - A.2 OAI-PMH: Erweiterte Anforderungen - A.2.1 Set für Open-Access-Dokumente: Dienste veröffentlichen neben Open-Access-Dokumenten oftmals auch Dokumente, die nicht frei verfügbar, sondern beispielsweise nur für einen institutionell eingeschränkten Nutzungskreis zugänglich sind. Zur Erbringung von Mehrwertdiensten durch Service Provider ist es wichtig, Datensätze, die sich auf Open-Access-Dokumente beziehen, von solchen zu unterscheiden, für die das nicht gilt, und darüber hinaus eine vorab durchgeführte Selektion zu erlauben. **Mindestanforderung:** M.A.2-1 Es existiert ein Set mit der Bezeichnung (setSpec) „open\_access“. Zu diesem Set gehören alle Metadatensätze, die sich auf Open-Access-Dokumente beziehen, d. h. bei diesen steht ein zugehöriger und verlinkter Volltext frei zur Verfügung.

<sup>9</sup> Mit dem Begriff *Frontdoor* wird eine, in der Regel dem Volltext vorgeschaltete Eingangsseite eines Dokuments im Angebot eines Repositoriums bezeichnet, die alle wichtigen Metadaten und den Volltextlink bereitstellt.

Die aus Sicht der Nutzer\*innen wichtigste Angabe, der Link zum Volltextdownload wird zuweilen durch umfangreiche Metadatenangaben verschleiert oder zu unauffällig in der Peripherie der Seite angeordnet.

### Probleme durch schwankende Verfügbarkeit einzelner Archive

Durch die extreme Vielfalt der Datenlieferanten folgt auch eine große Bandbreite in der Zuverlässigkeit der Archive. Schwankende Verfügbarkeitsquoten einzelner stark nachgefragter Repositorien erhöhen den Supportaufwand in der Beratung signifikant.

### Neue Dienste: Direktsupportoption am Dokument

Durch die besonderen Rahmenbedingungen der für Bibliotheken noch neuen Medienart der Open Access Texte ergeben sich auch Chancen für neue Formen des Bibliothekssupports bei der Katalognutzung. Die bereits geschilderten Nutzungshemmnisse einzelner Angebote führten an der Staats- und Universitätsbibliothek Bremen zur Entwicklung einer Direktsupportoption am jeweiligen Katalognachweis, um Fragen der Verfügbarkeit der Texte schnell und unkompliziert per Emailsupport von Bibliotheksmitarbeiter\*innen klären zu können. Dazu wird in jeder Dokumentvollanzeige eines BASE/OAI-Nachweises ein Emailformular direkt beim Volltextlink angeboten, über den eine Anfrage generiert werden kann, die automatisch die angefragten Metadaten des Titels mitliefert.

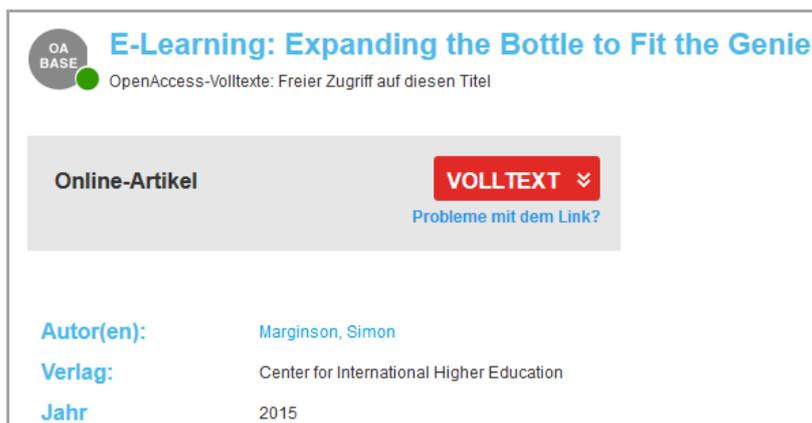


Abbildung 4: Direktsupport-Link in der Medienvollanzeige von Open Access Medien in der E-LIB Bremen: "Probleme mit dem Link?"

Durch den kurzfristigen Support können oft Wege eröffnet werden, das vollständige Dokument aus anderer Quelle zu beschaffen.

## Open Access Medien im Bibliothekskatalog / Technische Integration der Metadaten

### Filterung der BASE Metadaten und Suchmaschinenindexierung

Die BASE/OAI-Daten liegen mit knapp 85.000 Dateien in einer flachen Verzeichnisstruktur in einem erweiterten OAI-DC-Format<sup>10</sup> vor (180 GB unkomprimiert). Zur Identifikation der für den Import vorgesehenen Datensätze werden die Daten in Bezug auf die BASE-Verfügbarkeitsangabe im Tag "base\_dc:oa" analysiert (1: open access; 0: unknown; 2: no open access) und Archive des Types 2 ausgefiltert. Darüber hinaus werden jeweils eine an der SuUB Bremen gepflegte Negativ- und Positiv-Liste mit weiteren Einträgen von Repositorien berücksichtigt, um auch die speziellen Anforderungen des eigenen Standortes berücksichtigen zu können. Die Blacklist korrigiert nachträglich den Wert von "base\_dc:oa", so dass das Repository als "no open access" gewertet wird. Dagegen werden die Daten aller in der Whitelist enthaltenen Repositorien unabhängig vom "base\_dc:oa" als open access Metadaten gewertet und importiert.

Im April 2016 wurden von ca. 85 Millionen Datensätzen etwa 25 Millionen für den Import in die E-LIB identifiziert. Ca. 6 Millionen wurden durch die Blacklist ausgefiltert, alle übrigen nicht berücksichtigten Archive gehören zum Typ BASE "no open access". Die Whitelist betrifft aktuell nur einen sehr geringen Anteil aller Datensätze (160.000).

Dieser Filterprozess wurde mit der Skriptsprache Perl und – nach ersten Laufzeittests mit den Bibliotheken XML::Twig<sup>11</sup> und XML::XPath<sup>12</sup> – endgültig mit XML::LibXML<sup>13</sup> realisiert. Die Laufzeiten der Filterung unterscheiden sich bei dieser Datenmenge erheblich: So benötigte die erste Implementation auf Basis von XML::XPath ca. 80 Stunden, bei XML::Twig ca. 50 Stunden und mit XML::LibXML konnte die Laufzeit auf ca. 5 Stunden reduziert werden.

Im nächsten Schritt werden die Daten in den Index importiert. Alle verbliebenen relevanten 25 Mio. Metadatenätze aus den BASE/OAI-Daten werden mit Hilfe eines in ANSI-C geschriebenen, modular anpassbaren Importprogramms in 3 separate Index-Dateien<sup>14</sup> in einer Laufzeit von 9 Stunden importiert.

---

<sup>10</sup> OAI-DublinCore - <http://dublincore.org/documents/dces/> (Aufruf 21.04.16)

<sup>11</sup> XML::Twig <http://search.cpan.org/~mirod/XML-Twig-3.49/Twig.pm> (Aufruf 21.04.16)

<sup>12</sup> XML::XPath <http://search.cpan.org/~manwar/XML-XPath-1.36/lib/XML/XPath.pm> (Aufruf 21.04.16)

<sup>13</sup> XML::LibXML [http://search.cpan.org/dist/XML-LibXML/LibXML\\_pod](http://search.cpan.org/dist/XML-LibXML/LibXML_pod) (Aufruf 21.04.16)

<sup>14</sup> Die CiXbase Indizes liegen in einer flat-file-Struktur mit sog. B\*-Bäumen vor.

Die aufgebauten Indizes liegen – wie die jeweiligen Indizes anderer Datentypen – auf einem verteilten Cluster

aus 8 Servern – insgesamt etwa 60 Mio. Metadaten. Bei Suchanfragen werden alle verschiedenen Indizes parallel einzeln ausgewertet, anschließend gemischt, nach Relevanz sortiert und als ein aggregiertes Resultat über einen Frontend-Server an die Nutzer\*innen ausgeliefert.

### **Fazit**

Die Integration online frei verfügbarer wissenschaftlicher Texte besitzt großes Potential, das eigene Angebot um hochaktuelle, fremdsprachige aber auch historische Medien vielfältig zu erweitern und bisher wenig genutzte Quellen internationaler Dokumentenarchive in gleichzeitiger Sicht auf lokale Medien vor Ort verfügbar zu machen. Open Access Publikationen anderer Standorte werden damit an den Universitäten sichtbar und freie wissenschaftliche Publikationen genauso behandelt wie Zeitschriftenartikel aus klassischen Verlagsproduktionen.

Von besonderer strategischer Bedeutung ist zukünftig die zu verbessernde Nachweisqualität der Repositorienanbieter, um die Nachweise wirklich freier Dokumente von klassischen Verlagsmetadaten zu differenzieren.

Weitere Zukunftsprobleme stellen die meist noch fehlenden Verknüpfungen von Verlagsveröffentlichungen mit den potentiell vorhandenen parallelen Zweitveröffentlichungen auf freien Dokumentenservern und die fehlenden Verfügbarkeitsinformationen einzelner frei verfügbarer Open Accesspublikationen<sup>15</sup> in ansonsten lizenzpflichtigen (Hybrid-)Zeitschriften dar.

---

<sup>15</sup> Z.B. *OpenChoice* Publikationen beim Springer Verlag