

Marianna Mühlhölzer, Norbert Pfurr, Sven-S. Porst und Bernhard Wagenbreth

www.GEO-LEO.de – das Recherchewerkzeug des Fachinformationsdienstes für Wissenschaft zum System Erde und Weltall

Zusammenfassung: GEO-LEO ist ein zentrales Recherchewerkzeug zum System Erde und Weltall. Als integriertes Internetportal dient es zum Nachweis gedruckter und elektronischer Informationsressourcen der Sondersammelgebietsbibliotheken in Freiberg und Göttingen sowie anderer fachrelevanter Anbieter. Zudem bietet es der Fachgemeinschaft eine Publikationsplattform für digitale Dokumente im Sinne des Open Access an. Das Portal, welches auf moderner Open Source Software basiert, stellt eine gute Basis für eine Webpräsenz eines zukünftigen astro-, geo- und montanwissenschaftlichen Fachinformationsdienstes für Deutschland dar.

Schlüsselwörter: GEO-LEO; Fachportal; Metasuche; Publikationsserver; Astronomie; Geowissenschaften; Montanwissenschaften

www.GEO-LEO.de – Search Tool for Scientific Information in Earth and Space Sciences

Abstract: GEO-LEO provides an integrated search on printed and digital information resources of the special subject collections in the libraries at Freiberg and Göttingen as well as of other providers. In addition it includes an open access repository. GEO-LEO, using modern open source software, appears as a valuable tool for a website of a future Scientific Information Service in context of Astronomy, Geology, Mining and related subjects for Germany.

Keywords: GEO-LEO; subject gateway; meta search engine; Open Access repository; space sciences; earth sciences; geology; mining

Marianna Mühlhölzer: muehlhoelzer@sub.uni-goettingen.de
Dr. Norbert Pfurr: pfurr@sub.uni-goettingen.de
Sven-S. Porst: porst@sub.uni-goettingen.de
Bernhard Wagenbreth: Bernhard.Wagenbreth@ub.tu-freiberg.de

Inhalt

1	Einleitung	72
2	Portalaufbau	74
2.1	Suche	74

2.2	Thematische Suche	75
2.3	Elektronische Zeitschriften	75
2.4	Geo-Guide	75
2.5	GEO-LEO-e-docs	76
3	Technik	76
3.1	Pazpar2	77
3.1.1	Abfragesprachen	77
3.1.2	Abfrageprotokolle	77
3.1.3	Einsammeln der Ergebnisse	77
3.1.4	Normalisierung	77
3.1.5	Ausliefern der Ergebnisse	78
3.1.6	Konfiguration	78
3.1.7	Weiterentwicklung	78
3.2	OAI-Harvesting	78
3.3	TYPO3	79
3.3.1	Pazpar2 TYPO3-Extension	79
3.3.2	Nkwwgok-Themenbäume	80
3.3.3	EZB-Abfrage	80
3.3.4	Piwik	80
3.4	Open Source	80
3.5	DSPACE Dokumentenserver	80
3.5.1	Datenmodell	81
3.5.2	Benutzer- und Rechteverwaltung	81
3.5.3	Einstellen von Dokumenten und Geschäftsgänge	81
3.5.4	Persistente URL	81
3.5.5	Suche und Anzeige der Ressourcen	82
3.5.6	Neuerscheinungslisten	82
4	Ausblick	82

1 Einleitung

Geowissenschaftliche Forschung wird häufig als „kumulative Wissenschaft“ bezeichnet, da die Beobachtungen in Aufschlüssen von heute nicht die früherer Erforschungen ersetzen, zumal ältere Aufschlüsse heute meist gar nicht mehr zugänglich sind. Es ist daher üblich, bei aktuellen Forschungsvorhaben die Ergebnisse früherer Forschung zu berücksichtigen und über die Einbeziehung historischer Forschungsergebnisse in moderne Kontexte neue Erkenntnisse zu gewinnen. So kann die „Wiederent-

deckung“ von vergessenen Informationen zu bahnbrechenden Arbeiten in den Geowissenschaften führen. Zur Annäherung an diese – meist in Buchseiten in den Bibliotheken „verschollenen“ – Informationen kann „GEO-LEO“, das Rechercheportal zum System Erde und Weltall, wertvolle Hilfe bieten, da es das Auffinden von und Suchen bzw. Browsen in Informationsressourcen zum System Erde und Weltall ermöglicht. Inhaltlich reicht dabei die Bandbreite der Wissenschaften von Astronomie und Weltraumforschung über Geologie, Mineralogie und Geophysik bis hin zu angewandten Disziplinen wie Bergbau und Hüttenwesen.

Mitte der Neunziger Jahre setzte sich nach der Entwicklung breitenwirksamer und grafikfähiger Anwendungen des Internets die Erkenntnis durch, dass die in dieser Form entstandenen wissenschaftlichen, elektronischen Informationsträger für die Sondersammelgebiete (SSG) von großer Bedeutung sind. Daher regte die Deutsche Forschungsgemeinschaft (DFG) die SSG-Bibliotheken an, auch diese elektronischen Fachinformationen entsprechend dem bibliothekarischen Grundsatz¹ unter archivarischen, ökonomischen und synoptischen Gesichtspunkten zu sammeln, zu ordnen und verfügbar zu machen. Aus dem Anspruch der wissenschaftlichen Qualität ergab sich die Notwendigkeit, durch intellektuelle Recherche und Bewertung die Auswahl relevanter Fachinformationen in elektronischer Form zu gewinnen, diese formal zu beschreiben und inhaltlich zu erschließen, um sie dann in das Angebot der wissenschaftlichen Bibliotheken integrieren und dem Nutzer in strukturierter Form anbieten zu können. Diese neue Anforderung eines „Sondersammelgebiet-Fachinformationsführers“ (SSG-FI) wurde seit 1996 als Projekt „Geo-Guide“ exemplarisch für die geowissenschaftlichen Sondersammelgebiete an der Niedersächsischen Staats- und Universitätsbibliothek Göttingen (SUB Göttingen) umgesetzt² und seit 1999 als Gemeinschaftsprojekt mit der Universitätsbibliothek der TU Bergakademie Freiberg (UB Freiberg) weitergeführt³. So stellte der www-basierte Fachinformationsführer „Geo-Guide“ die Vorstufe der geo- und montanwissenschaftlichen Virtuellen Fachbibliothek „GEO-LEO“ dar, welche seit 2002 als neue Dienstleistung von der UB Freiberg und der SUB Göttingen entwickelt und betrieben wird und gedruckte

und elektronische Informationsressourcen zum System Erde und Weltall integriert. Seitdem stellt www.GEO-LEO.de das zentrale Recherchewerkzeug für geo- und montanwissenschaftliche Fachinformation für Deutschland dar, das auf Anregung der DFG 2009 mit der Einbeziehung von Astronomie und Weltraumforschung erweitert wurde. Das Portal führt die Angebote der Sondersammelgebietsbibliotheken, weiterer Kooperationspartner und der Nationallizenzen unter einer Suchoberfläche zusammen⁴.

Das in den letzten Jahren durch die „Open-Access-Initiative“ geforderte Prinzip des „offenen Zugangs“ zu wissenschaftlichen Erkenntnissen wurde insbesondere mit der Bereitstellung der Publikationsplattform für digitale Dokumente zum System Erde und Weltall „GEO-LEOe-docs“ sowie dem fachspezifischen Harvesten von Volltextservern in der GEO-LEO-Suche umgesetzt⁵.

Das erste „GEO-LEO“-System arbeitete auf der Grundlage von Open Source Software (LAMP-System) und umfangreichen, eigens programmierten Adaptionen, um die hohen Kosten kommerzieller Produkte bei der Anschaffung und Wartung zu vermeiden sowie eine Nachnutzung zu erleichtern⁶. Mit dem Serverbetrieb an der SUB Göttingen wurde in den letzten Jahren auch die Pflege der Software von dieser Bibliothek übernommen.

Der Nachteil des individuellen Pflegebedarfs bei knappen und wechselndem IT-Personal und die Heterogenität der Portalsoftware der an der SUB Göttingen installierten Virtuellen Fachbibliotheken führte zur Vereinheitlichung der technischen Infrastruktur aller dort betriebenen Portale mit moderner Technologie. Ein Grundbaustein ist dabei die Content-Management-Software TYPO3, die Inhalt, Struktur und Layout der Portalseiten trennt und die Bearbeitung und Pflege dieser einzelnen, sich jedoch gegenseitig aufeinander beziehenden Bestandteile in definierten Arbeitsabläufen ermöglicht. Während die Anlage und Pflege der Struktur und des Layouts der Portalseiten in den Händen der Entwickler der Abteilung Digitale Bibliothek der SUB Göttingen liegt, können die Inhalte (Texte, Bilder, Dateien) von Redakteuren, hier den verschiedenen Fachreferentinnen und Fachreferenten der an „GEO-LEO“ beteiligten Bibliotheken, direkt erstellt und gepflegt werden, ohne sich um technische Aspekte kümmern zu müssen. Die technischen Weiterentwicklungen des „GEO-LEO“-Portals inklusive des Moduls „Geo-Guide“ sowie des Publikationsservers „GEO-LEOe-docs“ wurden 2012 realisiert.

¹ In Anlehnung an die Beschreibung der Aufgaben von Bibliotheken bei: Ewert, Gisela; Umstätter, Walther: Die Definition der Bibliothek. In: Bibliotheksdienst 33 (1999), S. 966.

² Vgl. DBI-Materialien; 185 (1999): Schriften der Deutschen Forschungsgemeinschaft, 149 S.

³ Vgl. Freiburger Forschungshefte D 207 Montan- und Technikgeschichte, 2001, S. 315–324.

⁴ Vgl. Bibliothek 31 (2007), Nr. 3, S. 293–298.

⁵ Vgl. Bibliotheksdienst 40. Jg. (2006), H.10, S. 1115–1126.

⁶ Vgl. Bibliotheksdienst 39. Jg. (2005), H. 3, S. 335–345.

Mit der Einbeziehung des „Pangaea Data Publisher“, einem digitalen Bibliothekssystem für Daten aus der Erdsystemforschung und den Umweltwissenschaften, bietet „GEO-LEO“ seit einigen Monaten auch die übergreifende Suche bis hin zu wissenschaftlichen Forschungsdaten („Primärdaten“) an.

2 Portalaufbau

Seit April 2012 ist das Portal in einem gänzlich neuen Erscheinungsbild in den Weiten des WWW sichtbar. Für die Seiten wurde in Zusammenarbeit mit dem schwedischen Grafiker Henrik Cederblad ein modernes und leicht überschaubares Design entwickelt und gestaltet.

Die Grundstruktur der Seiten wird gegliedert durch eine Kopfzeile mit grafischen Elementen und einer Navigationsleiste, einen zentral positionierten, sich dynamisch aufbauenden inhaltlichen Teil (Funktionsseiten) sowie einer stets präsenten Fußzeile mit Links zu allgemeinen Informationstexten (Abbildung 1).

Die grafischen Elemente der Kopfzeile umfassen ein Hintergrundbild und das optimierte Logo von „GEO-LEO“. Ein schematisierter Profilschnitt eines fiktiven Geländeanschnitts, wie er von Geologen zur Darstellung von Gesteinsabfolgen und ihren Strukturen im Gelände üblich ist, sowie ein darüber ausgebreiteter Sternenhimmel mit einzelnen Spiralgalaxien dienen der Symbolisierung sowohl der Geo- und Montanwissenschaften als auch der Astronomie und Weltraumforschung. Die grafische Komplexität des ursprünglichen Logos wurde zugunsten einer besseren Lesbarkeit und schwarz-weiß-Reproduzierbarkeit reduziert und das modernisierte Logo mit klaren Kontrasten und linearer horizontaler Ausrichtung versehen.

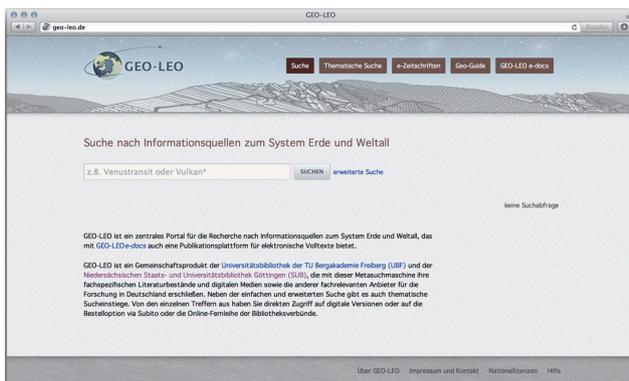


Abb. 1: Startseite des „GEO-LEO“-Portals.

Die verschiedenen Funktionsseiten sind über eine horizontal rechtsbündig in der Kopfzeile angelegte Navigations-

leiste ansteuerbar, wobei die rechteckigen Buttons auf die Menüpunkte *Suche*, *Thematische Suche*, *e-Zeitschriften*, *Geo-Guide* und *GEO-LEO-e-docs* verlinken.

Die Funktions- und Informationsseiten in „GEO-LEO“ werden über das TYPO3-CMS-Version 6 in einem Backend erstellt und über ein Frontend im Internet sichtbar.

2.1 Suche

Die erste Funktions- und gleichzeitig Einstiegsseite des Portals enthält das Modul Suche (Abbildung 1). Sie bietet derzeit Suchmöglichkeiten in folgenden fachlich relevanten Ressourcen:

Bibliothekskataloge

- GEO-OPAC – Sondersammelgebietskatalog Geowissenschaften
- Astro-OPAC – Sondersammelgebietskatalog Astronomie
- Katalog der UB der TU Bergakademie Freiberg und SWB⁷
- Göttinger Universitätskatalog und GVK⁸
- Katalog der ETH Zürich und NEBIS⁹
- Katalog der meereswissenschaftlichen Bibliotheken Deutschlands
- Katalog der Geographischen Zentralbibliothek im Leibniz-Institut für Länderkunde

Aufsatzdatenbanken

- Online Contents – SSG Geowissenschaften
- Online Contents – SSG Astronomie
- Agricola Article Citation Database

Alte Karten

- IKAR¹⁰ – Landkartendrucke vor 1850

Dokumentenserver

- Repositories
- Digitalisate

Datenanbieter

- PANGAEA Data Publisher for Earth and Environment

Internet-Quellen

- Geo-Guide

⁷ Südwestdeutscher Bibliotheksverbund.

⁸ Gemeinsamer Verbundkatalog.

⁹ Netzwerk von Bibliotheken und Informationsstellen in der Schweiz.

¹⁰ Altkartendatenbank.

Wie aus den ersten Versionen des Portals gewohnt, kann der Nutzer gleichzeitig über alle eingebundenen Angebote suchen; das Konzept einer „One-Stop-Recherche“ ist damit umgesetzt. Neben der einfachen Suche über alle Felder in einem Suchschlitz kann man auch auf eine erweiterte Suche mit vorgegebenen Suchfeldern umschalten und so gezieltere Suchen mit kleineren Treffermengen ausführen. Neu ist die Möglichkeit, nach Begriffen in den im Rahmen der Kataloganreicherung eingebundenen Inhaltsverzeichnissen zu einzelnen Publikationen zu suchen. Weitere Eingrenzungen der Treffermengen können komfortabel über mit der jeweiligen Trefferliste individuell erzeugte Facetten am rechten Seitenrand vorgenommen werden (siehe Abschnitt Pazpar2, Abbildungen 2 und 7). Insgesamt ist das Layout der erweiterten Suche gegenüber der älteren Version übersichtlicher gestaltet.

2.2 Thematische Suche

Die zweite Funktionsseite beinhaltet eine thematische Suche. Sie bietet dem Nutzer eine hohe Trefferpräzision, da sie auf die systematische Erschließung der Ressourcen zurückgreift. Der für das Browsing angebotene Themenbaum basiert auf der „Göttinger Online-Klassifikation“ (GOK), die als Normdatensatz vorliegt und über ein Mapping mit der in Freiberg verwendeten Dezimalklassifikation verknüpft ist. Über sieben Einstiegsthemen mit derzeit knapp 5.000 geowissenschaftlichen, 300 montanwissenschaftlichen und 350 astronomischen Themengebieten gelangt der Nutzer zielgenau und verbundübergreifend zu den Bibliotheksbeständen der Sondersammelgebiete in Freiberg und Göttingen sowie zu den im Fachinformationsführer „Geo-Guide“ aufgenommenen Internetressourcen (vgl. u.).



Abb. 2: Thematische Suche im Segment Geophysik.

Nach der Wahl eines Fachgebiets werden alle zugehörigen Treffer angezeigt. Gleichzeitig öffnet sich eine Auswahlbox, in der das ausgewählte Fachgebiet thematisch untergliedert wird. Durch Anklicken eines dieser untergeordneten Themen startet eine neue Suche. Dieser Vorgang kann so lange wiederholt werden, bis die letzte Stufe der Fachgliederung erreicht ist (Abbildung 2). Daneben kann der Nutzer die Treffermengen durch die Eingabe eines weiteren Suchbegriffs in einem separaten Suchfeld eingrenzen. Die Thematische Suche wird an dieser Stelle mit einer Suche über alle Felder kombiniert. Auch die Gestaltung dieser Seite wurde umfassend überarbeitet.

2.3 Elektronische Zeitschriften

Die dritte Funktionsseite bietet den fachrelevanten Ausschnitt der „Elektronischen Zeitschriftenbibliothek“ (EZB) im Look and Feel des „GEO-LEO“-Portals. Der Nutzer kann sich alle E-Zeitschriften der einzelnen Fachgebiete anzeigen lassen und über einen Suchschlitz nach einer bestimmten E-Zeitschrift suchen. Durch eine automatisierte Zuordnung der IP-Adresse des Nutzers zu einer Institution werden die jeweiligen Zugriffsrechte seines Standorts in der EZB-typischen Ampelfunktion angezeigt. Zu jedem Zeitschriftentitel sind die in der EZB enthaltenen Daten inklusive Link zum direkten Zugriff auf die Zeitschrift angehängt. Auch bei dieser Seite wurde die Usability durch ein überarbeitetes Layout verbessert.

2.4 Geo-Guide

Die vierte Funktionsseite bietet eine separate Sicht auf den „Geo-Guide“. Dieser Fachkatalog wissenschaftlich relevanter Internetquellen zum System Erde und Weltall wird von den Bibliotheken in Göttingen und Freiberg herausgegeben. Nachdem der Datenbestand 2012 überarbeitet und von Redundanzen z.B. zu den elektronischen Zeitschriften bereinigt wurde, erschließt der Katalog derzeit über 3.200 Webseiten (Stand Juli 2013). Wie bei der thematischen Suche kann sich der Nutzer alle Webseiten zu einem gesamten und zu untergliederten Themengebieten anzeigen lassen und bei Bedarf die Ergebnisse über die Facetten Quellenart, Land und Sprache gezielt filtern (Abbildung 3). Außerdem ist eine freie Suche über einen Suchschlitz möglich.

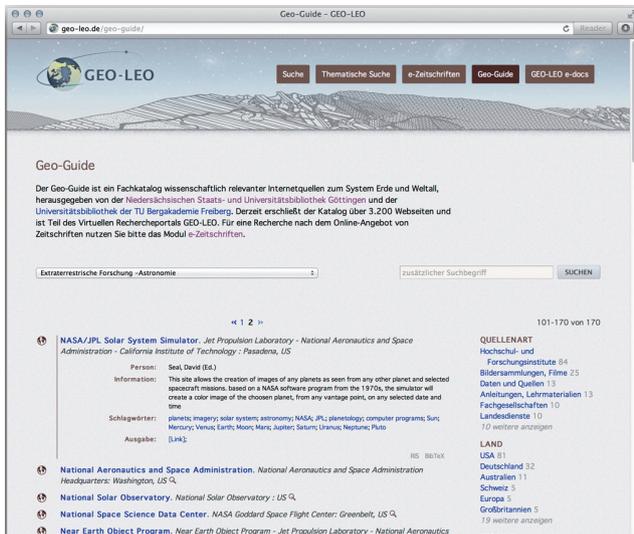


Abb. 3: Trefferanzeige im Geo-Guide.

2.5 GEO-LEO-e-docs

Die fünfte Funktionsseite verlinkt auf „GEO-LEO e-docs“¹¹, den Open-Access-Publikationsserver, der im Rahmen des von 2004 bis 2007 geförderten DFG-Projekts zum Aufbau der Virtuellen Fachbibliothek „GEO-LEO“ als zentrale Publikationsplattform eingerichtet wurde¹². Mit Projektende ging auch der Publikationsserver in den Routinebetrieb der SUB Göttingen über.

Derzeit ist auf dem Publikationsserver eine Sammlung von über 2.150 frei zugänglichen, digitalen Volltexten aus den Wissenschaftsgebieten der Göttinger und Freiburger Sondersammelschwerpunkte archiviert. Zum einen sammeln die beiden Bibliotheken thematisch relevante Volltexte, zum anderen können Wissenschaftlerinnen und Wissenschaftler selbst ihre Publikationen über „GEO-LEO e-docs“ open access zugänglich machen. Darüber hinaus sind Fachgesellschaften und wissenschaftliche Einrichtungen eingeladen, eigene Sammlungen einzurichten, auf denen sie selbstständig ihre Serien, Zeitschriften und andere Informationsressourcen veröffentlichen können. Das „GEO-LEO“-Team bietet dafür unterschiedliche Formen der Unterstützung und Zusammenarbeit an. Eine Kooperation besteht beispielsweise mit der „Geozon Science Media UG“¹³. Als Ausgründungsprojekt des Instituts für Geographie und Geologie der Universität Greifswald und eines Open Access-Verlags für Geo- und Umweltwissenschaften stellt „Geozon“ seine

11 <http://e-docs.geo-leo.de/>.

12 Vgl. Bibliotheksdienst 40. Jg. (2006), H.10, S. 1115–1126.

13 <http://www.geozon.info/>.

gesamten Publikationen in „GEO-LEO e-docs“ ein. Für derartige automatisierte Einspielungen großer Dokumentpakete stellt das System eine entsprechende Schnittstelle zur Verfügung (siehe Abschnitt DSpace Dokumentenserver).

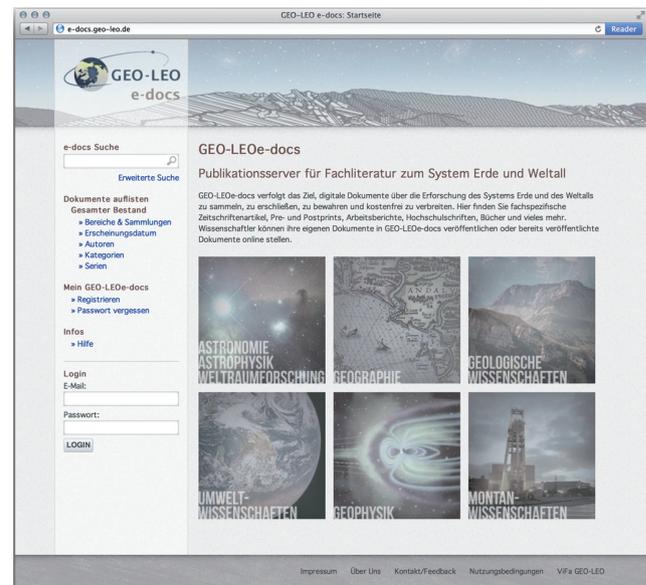


Abb. 4: Startseite „GEO-LEO e-docs“

Im Jahr 2012 wurde die dem Publikationsserver zugrunde liegende Software DSpace neu aufgesetzt. Mit diesem Schritt waren grundlegende konzeptionelle Veränderungen verbunden (Abbildung 4):

- Neu strukturiert orientieren sich die verschiedenen Fachbereiche an den überregionalen Sammelschwerpunkten der beteiligten Bibliotheken.
- Hinter den thematischen Bereichen verbergen sich nach Medientypen gegliederte Sammlungen.

Neben diesen konzeptionellen Änderungen hat sich auch das Erscheinungsbild des Publikationsservers maßgeblich geändert. Unter Mitwirkung des Grafikers Henrik Cederblad wurde das Layout modernisiert und dem des Rechercheportals angepasst (Abbildung 4).

3 Technik

Die Umsetzung des „GEO-LEO“-Portals beruht auf mehreren Technologien, die im Folgenden vorgestellt werden:

- Metasuch-Service pazpar2
- OAI-Harvesting
- Content-Management-System TYPO3 mit Erweiterungen

– DSpace Dokumentenserver
Die Auswahl dieser technischen Lösungen erfolgte mit Blick auf die Wartbarkeit und das Vorhandensein von Know-how zu der jeweiligen Software. Sämtliche Softwarekomponenten sind an der SUB Göttingen auch zu anderen Zwecken (andere Fachportale und Dokumentenserver) im Einsatz und unter freien Lizenzen verfügbar.

3.1 Pazpar2

Pazpar2 ist eine Open-Source-Metasuchsoftware von der Firma Index Data¹⁴. Mit ihr können Datenquellen über die bibliographischen Schnittstellen Z39.50 und SRU abgefragt werden. Darüber hinaus ist auch eine direkte Abfrage von Solr-Indizes möglich.

Pazpar2 ist ein reiner Datendienst: Er stellt Antworten auf Suchanfragen als Daten zur Verfügung. Die Nutzeroberfläche zur Darstellung der Ergebnisse ist ein separates Modul, wodurch der Metasuchdienst unabhängig von der darunterliegenden Technik und dem Webseitendesign betrieben werden kann. Dank einer Schnittstelle über das Web-Protokoll http lässt sich pazpar2 gut in Webseiten einbinden und ermöglicht eine im Vergleich zu älteren Lösungen sehr schnelle und dynamische Anzeige der Rechercheergebnisse, da sie eingehende Recherchetreffer der Metasuche umgehend ausliefert.

Die Arbeit von pazpar2 besteht aus mehreren Phasen:

1. Umwandeln der Nutzeranfrage in Abfragesprachen der Zieldatenbanken
2. Stellen der Anfragen an die Zieldatenbanken mit dem jeweils passenden Protokoll
3. Einsammeln der Ergebnisse von den Zieldatenbanken
4. Normalisieren der Ergebnisse
5. Ausliefern der aktuellen Ergebnismenge

3.1.1 Abfragesprachen

Pazpar2 akzeptiert strukturierte Abfragen in der *Common Command Language* (CCL, ISO 8777). In dieser Sprache lassen sich Boolesche Abfragen von verschiedenen Feldern ausdrücken, z. B. die Abfrage „Grönland and person = "Wegener, A*"“ für eine Suche nach „Grönland“ in allen Feldern und rechtstrunkiert nach der Phrase „Wegener, A“ im Feld „person“.

Die hierbei verfügbaren Feldnamen werden in der pazpar2-Konfiguration eingestellt und auf die Feldnamen

bzw. -nummern der jeweiligen Zieldatenbank abgebildet. Die Boolesche Struktur der Abfrage wird in die für die Zieldatenbank nötige Abfragesprache übersetzt. Die Beispielabfrage wird für eine SRU-Abfrage im CCL Format belassen, die Feldnamen aber werden angepasst: „Grönland AND pica.per: "Wegener, A*"“. Für eine „Z39.50“-Abfrage wird sie in die Prefix Query Format (PQF)-Abfrage „@and @attr 1 = 1016 Grönland @attr 1=1003 @attr 5 = 1 "Wegener, A*"“ mit den entsprechenden Bib-1-Attributen konvertiert. Für Solr-Indizes wird die Solr-Abfrage „Grönland AND pers: "Wegener, A*"“ mit den für diese Schnittstelle konfigurierten Indexnamen verwendet.

3.1.2 Abfrageprotokolle

Pazpar2 kann Abfragen über die Z39.50- und SRU-Protokolle sowie über die http-Schnittstelle von Solr ausführen. Die nötigen Details wie Protokollversionen und Zeichencodierungen sind hierbei wenn nötig konfigurierbar.

3.1.3 Einsammeln der Ergebnisse

Pazpar2 sammelt die Rechercheergebnisse über das jeweilige Protokoll in kleinen Mengen à 20 Datensätze ein. Hierdurch sind die ersten Treffer fast sofort verfügbar und werden dem Nutzer umgehend angezeigt.

Die Ergebnisse müssen als XML oder im ISO 2709-MARC-Format vorliegen, um von pazpar2 bearbeitet zu werden. Unstrukturierten Text (SUTRS) oder Binärformate wie MAB2 kann die Software dagegen nicht lesen.

3.1.4 Normalisierung

Der wichtigste Schritt der Datenverarbeitung durch pazpar2 ist die Normalisierung. Hierfür werden durch die Konfiguration Felder eines internen Datenmodells definiert, in die die Felder der geladenen Daten abgebildet werden müssen. Für Daten im MARC-Format liefert die pazpar2-Installation mit der *tmarc.xsl*-Datei eine Standardkonversion, die viele Felder extrahiert. So lässt sich mit wenig Aufwand eine gute Einbindung von MARC-Daten erreichen.

Um reichhaltigere Daten für die Anzeige zur Verfügung zu stellen, hat die SUB Göttingen diese Konversion erweitert, um spezielle Felder aus dem Bereich Geographie (Koordinaten von Landkarten) oder dem deutschen Bibliothekswesen (ZDB-Nummer) zu berücksichtigen. Eine Normalisierung der verfügbaren Medientypen auf eine überschaubare Anzahl und der Sprachcodes auf den ISO 639–

¹⁴ <http://www.indexdata.com/pazpar2>.

2/B-Standard ermöglicht das Filtern der Rechercheergebnisse nach diesen vorgegebenen Facetten.

Die Normalisierung funktioniert als XML/XSL-Workflow: Daten werden als XML eingelesen bzw. aus dem ISO 2709-MARC-Format automatisch in XML umgewandelt. Für jede Zieldatenbank wird eine Folge von XSL-Dateien festgelegt, die auf jeden Datensatz angewandt wird. So können wiederkehrende Konversionsschritte als einzelne XSL-Dateien abgelegt und nach Bedarf kombiniert werden.

Felder des internen Datenmodells können zur Nutzung für die Deduplizierung markiert werden: pazpar2 fügt dann alle Datensätze zusammen, bei denen der Inhalt dieser Felder übereinstimmt. So können – eine einigermaßen homogene Katalogisierung vorausgesetzt – gleiche Treffer für denselben Titel, verschiedene Erscheinungsformen (z.B. elektronisch und gedruckt) oder verschiedene Auflagen desselben Buches zu einem Treffer in der Ergebnisliste zusammengeführt werden. Die Detailinformation aus den verschiedenen Datenquellen bleibt hierbei erhalten und kann in die Darstellung einfließen.

3.1.5 Ausliefern der Ergebnisse

Die Kommunikation mit pazpar2 geschieht über das http-Protokoll. Sobald eine Suche gestartet ist, ist ein Abruf der bereits geladenen Treffer möglich. Damit können erste Ergebnisse der Suche umgehend angezeigt werden.

3.1.6 Konfiguration

Die pazpar2-Konfiguration liegt in XML-Dateien. In ihnen werden für den pazpar2-Server verschiedene Services eingerichtet, wobei jeder dieser Services eine Sammlung von Datenquellen als Metasuche konfiguriert. Im Fall von „GEO-LEO“ gibt es einen Service mit den oben genannten Quellen für die Metasuche auf der Startseite und einen weiteren nur mit den Beständen aus Freiberg und Göttingen für die thematische Suche. Die Konfiguration des verwendeten Protokolls, der Suchfelder und der Normalisierung wird für jede Datenquelle einzeln vorgenommen und kann in eigenen Dateien abgelegt werden. So lassen sich die Konfigurationen für häufig genutzte Server leicht wiederverwenden.

3.1.7 Weiterentwicklung

Pazpar2 wird von der Firma Index Data aktiv entwickelt. Gemeldete Fehler wurden häufig schnell von der Firma

behooben. Über die „yazlist“-Mailingliste können Fragen zur Softwarekonfiguration mit den Entwicklern und anderen Nutzern diskutiert und geklärt werden.

3.2 OAI-Harvesting

Die in der „GEO-LEO“-Metasuche eingebundenen Datenbestände „Digitalisate“, „Repositories“ und „Pangaea Data Publisher“ enthalten über OAI-PMH geharvestete Metadaten von über 100 Dokumentenservern aus aller Welt.

Der verwendete Harvester wurde von Timo Schleier für EROMM¹⁵ entwickelt und zeichnet sich durch eine Web-Oberfläche zur Konfiguration aus. Mit der grafischen Oberfläche können die Fachreferentinnen und Fachreferenten direkt die inhaltliche Auswahl der zu harvestenden Daten treffen und sofort konfigurieren.

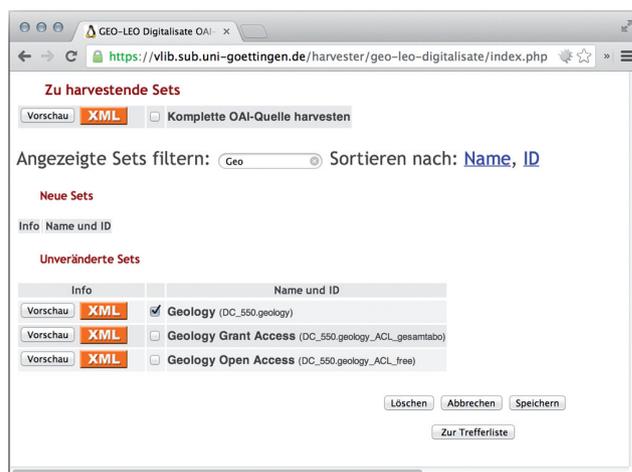


Abb. 5: Filtern und Auswahl von Sets in der Web-Oberfläche des Harvesters.

Nach Eingabe der URL einer OAI-PMH-Schnittstelle lädt der Harvester die vom Server angebotenen Sets und zeigt sie zur Auswahl an. Sets, die seit dem letzten Bearbeiten hinzugekommen oder verschwunden sind, werden dabei extra hervorgehoben. Eine Liste aktueller Fehlermeldungen in der Web-Oberfläche vereinfacht es, Probleme wie unauffindbare Server (häufig geänderte Adressen der OAI-Schnittstellen) zu erkennen und zu beheben. Im Hintergrund laufen regelmäßig Skripte, die den Datenbestand von den konfigurierten OAI-Servern aktualisieren und die Daten in Solr-Indizes einbinden. Die Solr-Indizes werden für die Metasuche direkt von pazpar2 abgefragt.

¹⁵ European Register of Microform and Digital Masters: <http://www.eromm.org>.

3.3 TYPO3

Der TYPO3-Server der SUB Göttingen stellt eine fertige Umgebung für TYPO3-Installationen zur Verfügung. Auf diesem laufen u. a. die drei an der SUB Göttingen betriebenen virtuellen Fachbibliotheken „GEO-LEO“, „Library of Anglo-American Culture“¹⁶ und „vifamath“¹⁷. Zur Präsentation der speziell in den virtuellen Fachbibliotheken vorhanden Inhalte gibt es TYPO3-Extensions:

- pazpar2: Einbindung der Metasuche
- nkwgok: Einbindung fachlicher Themenbäume
- ezbrequest: Darstellung von EZB-Daten

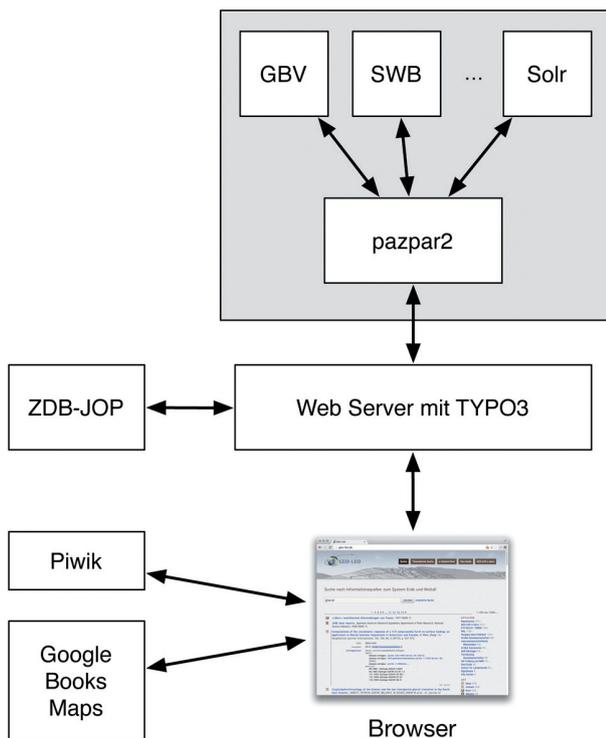


Abb. 6: Struktur und Abhängigkeiten der verschiedenen an GEO-LEO beteiligten Dienste.

3.3.1 Pazpar2 TYPO3-Extension

Die *pazpar2*-TYPO3-Extension stellt die Nutzeroberfläche für die Metasuche zur Verfügung. Sie kümmert sich um die Anzeige des Suchformulars auf der Webseite und enthält die JavaScript-Client-Software, die sich um die Anzeige der Ergebnisse und die Interaktion mit dem Nutzer kümmert.

16 <http://aac.sub.uni-goettingen.de>.

17 <http://vifamath.de>.

Die Funktionen der Client-Software sind umfangreich: Sie beinhalten die Anzeige der Ergebnisliste mit Blättern und Facettierung und die möglichst detailreiche Darstellung der Trefferdaten. Die Detailanzeige beginnt mit den bibliographischen Metadaten wie Titel, Personen, Serien und Zeitschriften und reicht bis zu speziellen Features wie der Verfügbarkeitsanzeige von Zeitschriften für die IP-Adresse des Nutzers über den Journals-Online-&Print-Dienst der ZDB (ZDB-JOP)¹⁸, der Einbindung von Google Books, der Einbindung von Google Maps für Karten mit Geoinformationen in den Metadaten, dem Export der Metadaten im RIS- oder BibTeX-Format für Literaturverwaltungsprogramme oder der Suche nach dem Titel im Karlsruher Virtuellen Katalog (KVK)¹⁹.

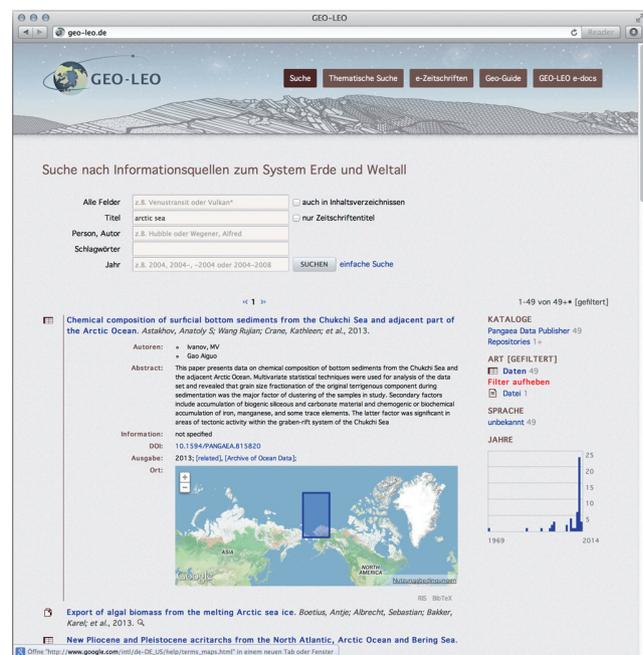


Abb. 7: Trefferanzeige inklusive einer Einbindung eines Kartenausschnitts zum Untersuchungsgebiet in Google Maps.

Die Umsetzung einer Pazpar2 Suche für vifanord²⁰ mit Einbindung in TYPO3 ist in Arbeit und die Anzeigelogik dieser TYPO3 Extension wird ebenfalls in der Metasuche der Virtuellen Fachbibliothek Ost- und Südostasien „CrossAsia“²¹ genutzt.

18 <http://www.zeitschriftendatenbank.de/services/journals-online-print>.

19 <http://www.ubka.uni-karlsruhe.de/kvk.html>.

20 <http://www.vifanord.de>.

21 <http://crossasia.org>.

3.3.2 Nkwgok-Themenbäume

Zur Darstellung von Themenbäumen existiert die TYPO3-Extension *nkwgok*. Sie lädt Informationen zur Struktur der Themenbäume aus einfachen CSV-Dateien und alternativ aus lokalen Normdatensätzen im ica OPAC.

Die CSV-Dateien enthalten für jedes Themengebiet eine Zeile mit jeweils der ID des Datensatzes, dem deutschen und englischen Namen des Themas, der ID des Elternelements und einer zugehörigen Suchabfrage. Für die „GEO-LEO“-Themensuche haben die zuständigen Fachreferentinnen und Fachreferenten diese CSV-Dateien erstellt und mit teilweise komplexen gleichzeitigen Suchabfragen in der „Göttinger Online Klassifikation“ und „Freiberger Dezimalklassifikation“ ausgestattet. Für die Suche im „Geo-Guide“ existiert ein weniger granularer Baum.

Bei der Einbindung in die Webseite löst eine Themenauswahl sofort die pazpar2-Suche über den jeweiligen Datenbestand aus. Die beiden TYPO3-Extensions sind an dieser Stelle gekoppelt.

Eine weitere Kopplung der beiden Extensions ist bei der Nutzung zur Anzeige von Neuerwerbungen möglich, die in „GEO-LEO“ momentan nicht genutzt, aber in den weiteren von der SUB Göttingen betriebenen Virtuellen Fachbibliotheken „Library of Anglo-American Culture“, „vifamath“ und den Webseiten der SUB Göttingen verwendet wird. Hier kann ein grober Themenbaum mit Abfragen definiert werden, und es wird für jedes Thema eine Checkbox angezeigt. Nutzer können sich dann die Neuerwerbungen zu diesem Thema anzeigen lassen oder sie als RSS Feed abonnieren.

3.3.3 EZB-Abfrage

Zur Anzeige der elektronischen Zeitschriften ist die Extension *ezbrequest* im Einsatz, die die XML-Schnittstelle der EZB nutzt und für „vifamath“ entwickelt wurde. TYPO3-Extensions mit ähnlicher Funktion gibt es auch an der SUB Hamburg (*libconnect*), an der UB der Humboldt-Universität zu Berlin (*huubzeitschriftendienst: evifa*) und der BSB München (*bsbezb: vifaost*).

3.3.4 Piwik

Neben den eigens entwickelten Extensions kommt die bereits existierende Extension *piwik* zum Einsatz, die für die Zählung aller Seitenaufrufe des TYPO3-Systems mit der Nutzerzählungssoftware Piwik²² sorgt. Durch den Einsatz derselben modernen Zählsoftware auf allen Webservern der SUB Göttingen erhält man eine realistische und konsistente Nutzungsstatistik.

3.4 Open Source

Da sowohl der Aufbau als auch der Betrieb von GEO-LEO durch öffentliche Mittel finanziert werden, sind auch alle in diesem Rahmen entstandenen Programme und Konfigurationen frei verfü- und weiterverwendbar. Sie sind wie auch die Entwicklungen von anderen Projekten der SUB Göttingen im Account *subugoe* bei GitHub²³ abgelegt. Insbesondere:

- pazpar2-Konfiguration: <https://github.com/subugoe/pazpar2-SUB>
- OAI-PMH-Harvester: <https://github.com/subugoe/SUB-Harvester>
- TYPO3-Konfiguration für GEO-LEO: <https://github.com/subugoe/vlibs-typo3/tree/geoleo>

3.5 DSpace Dokumentenserver

Die SUB Göttingen hat sich nach der erfolgreichen Entwicklung von „GEO-LEO e-docs“ unter der freien Software DSpace²⁴ für deren Einsatz zum Aufbau ihrer gesamten Dokumentenserverarchitektur entschieden. Im Rahmen eines Gesamtkonzepts für die Erfassung, Speicherung und Weiterverbreitung von digitalen Ressourcen der Bibliothek für die Universität Göttingen (z. B. Dissertationen, Universitätsverlag, Akademie der Wissenschaften u. a.) profitiert so auch „GEO-LEO e-docs“ von der in diesem Rahmen vorgenommenen Anpassung und Weiterentwicklung der Werkzeuge, die die Software zur Verfügung stellt. Eine Nachnutzung des Konzepts für andere Fachportale im Kontext der Virtuellen Fachbibliotheken ist in Planung.

Das System läuft unter einer JAVA-Laufzeitumgebung Version 1.60_26 und dem Betriebssystem LINUX. Es ver-

²² <http://de.piwik.org>.

²³ <https://github.com/subugoe>.

²⁴ <http://www.dspace.org/>.

wendet eine PostgreSQL-Datenbank und hat eine 3-Schicht-Architektur:

- In einer Anwendungsschicht sind die Komponenten enthalten, mit denen „GEO-LEO e-docs“ mit der Außenwelt kommuniziert (Webschnittstelle, OAI-PMH- und SWORD-Schnittstelle).
- Eine Geschäftslogik dient zur Verwaltung von Inhalt, Benutzern und Gruppen, Rechten und Geschäftsgängen.
- In der dritten Schicht werden die Metadaten und Inhalte abgespeichert.

3.5.1 Datenmodell

Das Datenmodell ist nach einem Container-Prinzip organisiert:

Bereiche -> Teilbereiche -> Sammlungen -> Dokument -> Metadaten + Datei

Die Dateien sind in Bündeln organisiert, die die ursprünglichen Dokumente, Vorschauminiaturen von Bildern zur Anzeige in Übersichten, extrahierten Text aus den Dokumenten sowie die Lizenzbestimmungen, unter denen eine Ressource veröffentlicht wurde, enthalten.

Die Metadaten zu jedem Dokument sind deskriptive, administrative und strukturelle Metainformationen, die in einer Metadatenreferenzliste zusammengefasst sind. Prinzipiell können verschiedene Metadatenschemata eingebunden werden. In „GEO-LEO e-docs“ wird das Dublin-Core-Schema genutzt.

3.5.2 Benutzer- und Rechteverwaltung

Die Benutzer- und Rechteverwaltung regelt das Rechtmanagement für verschiedene Rollen und deren Aufgaben in „GEO-LEOe-docs“.

Ein Benutzer registriert sich mit seiner E-Mail-Adresse und einem Passwort. Damit gehört er zur Gruppe der Autoren und ist berechtigt, auf „GEO-LEOe-docs“ zu publizieren.

Daneben gibt es noch Editoren und Administratoren für die verschiedenen thematischen Bereiche.

Editoren sind für den Inhalt ihres Bereichs verantwortlich. Sie überprüfen eingereichte Publikationen, können sie ablehnen oder freischalten und deskriptive Daten ändern.

Administratoren sind für die Verwaltung ihres Bereiches zuständig. Dies umfasst das Löschen und Anlegen von Sammlungen, das Management der Zugriffsrechte auf Publikationen und zugehörige Dokumente sowie der Gruppenzugehörigkeit von Benutzern.

Durch diese klare Trennung der Rollen haben Autoren, Administratoren und Editoren jeweils eigene Sichten, mit denen sie Sammlungen bzw. Dokumente verwalten können. Die Kommunikation zwischen den drei Rollen findet mit einem integrierten E-Mail-System statt. Derzeit sind 450 Benutzer registriert.

3.5.3 Einstellen von Dokumenten und Geschäftsgänge

Dokumente können einzeln über das Web-User-Interface oder per Stapelverarbeitung über die SWORD-Schnittstelle eingespielt werden. In der Regel durchlaufen die Dokumente einen Geschäftsgang, in dem mehrere Bearbeitungsschritte definiert werden können. Deskriptive Metadaten bereits veröffentlichter Dokumente können über einen Identifier (PPN oder DOI) aus der Pica-Datenbank des GBV bzw. von CrossRef importiert werden (Abbildung 8). Im Gegenzug existiert ein Werkzeug, mit dem bibliographische Daten von Publikationen in die Pica-Datenbank exportiert werden können.

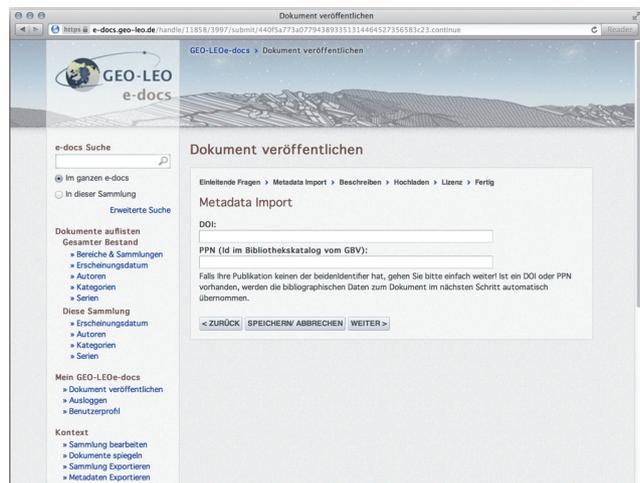


Abb. 8: Eingabeformular zur Generierung eines Metadatenimports aus der PICA-Datenbank bzw. CrossRef.

Obwohl „GEO-LEOe-docs“ beliebige Dokumentformate (PDF, word, excel, rtf, postscript etc.) zur Einstellung akzeptiert, wird eine Publikation im PDF- bzw. PFD/A-Format präferiert. Nur so ist garantiert, dass eine Publikation über einen längeren Zeitraum nutzbar ist.

3.5.4 Persistente URL

Für jede Publikation in „GEO-LEO e-docs“ (Bereiche, Teilbereiche, Sammlungen, Bündel und Dateien) werden per-

sistente Identifier unter Anwendung des CNRI-Handle-System vergeben.

3.5.5 Suche und Anzeige der Ressourcen

Die Inhalte von „GEO-LEOe-docs“ werden dem Nutzer durch verschiedene Funktionalitäten erschlossen: Zum einen kann man sich den Bestand an Dokumenten über die thematischen Bereiche und Sammlungen nach Erscheinungsdatum, Autoren, Serien und Kategorien auflisten lassen und in den erzeugten Listen browsen. Dazu werden durch die Suchmaschine Lucene Indices aus den Metadaten erzeugt. Die Auflistung über Kategorien nutzt beispielsweise die Indices der vergebenen Klassenbezeichnungen der GOK und Schlagworte.

Zum anderen stehen eine einfache Stichwortsuche und eine erweiterte Suche mittels Solr zur Verfügung. Dabei werden die aus den Metadaten und den Volltexten extrahierten Textelemente durchsucht.

3.5.6 Neuerscheinungslisten

Um sich regelmäßig über die Neuerscheinungen zu informieren, kann der Leser RSS-Feeds zu den verschiedenen Sammlungen abonnieren.

4 Ausblick

Die Basis von „GEO – Library Experts Online“, dem Researchwerkzeug nach Informationsquellen zum System Erde und Weltall, stellen die in den Bibliotheken vorhandenen und Dank der DFG-Förderung ständig erweiterten fachwissenschaftlichen Bestände dar. Unverzichtbare Datengrundlage für die Suche bilden die mit inhaltsbeschreibenden standardisierten Deskriptoren angereicherten Literaturdatenbanken, geschaffen durch jahrzehntelange fachwissenschaftliche Erschließungsarbeit vieler Fachreferentinnen und Fachreferenten.

Die Anwendung angemessener Informationsfilter bzw. Sortierungsmöglichkeiten erfordert bei der weiterhin anhaltenden „Informationsexplosion“ qualitativ hochwertige Metadaten, um die von den Nutzern erwartete Präzision der Rechercheergebnisse realisieren zu können. Neuere Entwicklungen versuchen, automatisiert normierte und strukturierte Metadaten mit Hilfe „lernfähiger“ Algorithmen über Zusammenhänge innerhalb der sprachlichen Beschaffenheit der Begriffswelt zu erzeugen. Die Zukunft wird zeigen, inwieweit es möglich ist, die zwar personal-

intensive, aber qualitativ hochwertige intellektuelle inhaltliche Erschließungsarbeit durch derartige automatisierte Verfahren zu ersetzen.

Die im Vergleich zu anderen Virtuellen Fachbibliotheken über Jahre anhaltend guten Zugriffszahlen zeigen, dass „GEO-LEO“ einen stabilen Nutzerkreis geo- und montanwissenschaftlicher Wissenschaftlerinnen und Wissenschaftler aufweist.

Derart etabliert und mit moderner, leistungsstarker Informationstechnologie ausgerüstet, stellt „GEO-LEO“ eine gute Ausgangsbasis für eine Webpräsenz der zukünftigen Fachinformationsdienste zum System Erde und Weltall dar, die entsprechend den Anforderungen als leistungsfähiges Informationssystem für die Wissenschaft unter überregionalen Gesichtspunkten weiterentwickelt werden wird. Dies entspricht den Richtlinien „Fachinformationsdienste für die Wissenschaft“ der Deutschen Forschungsgemeinschaft vom März 2013²⁵, nach denen die Aufgabe des Fachinformationsdienstes neben der qualifizierten und bedarfsgerechten Informationsversorgung der jeweiligen Fachcommunity durch die überregionale Bereitstellung auch im Aufbau und der Pflege komfortabler Nachweis- und Recherchesysteme besteht, welche technisch auf dem neuesten Stand gehalten werden sollen.

Marianna Mühlhölzer

Abteilung Digitale Bibliothek
Niedersächsische Staats- und Universitätsbibliothek Göttingen
D-37070 Göttingen
muehlhoelzer@sub.uni-goettingen.de

Norbert Pfurr

Fachreferat Geowissenschaften und Astronomie
Niedersächsische Staats- und Universitätsbibliothek Göttingen
D-37070 Göttingen
pfurr@sub.uni-goettingen.de

Sven-S. Porst

Abteilung Digitale Bibliothek
Niedersächsische Staats- und Universitätsbibliothek Göttingen
D-37070 Göttingen
porst@sub.uni-goettingen.de

Bernhard Wagenbreth

Fachreferat Geotechnik, Bergbau, Markscheidewesen
Universitätsbibliothek der TU Bergakademie Freiberg
Agricolastraße 10
D-09599 Freiberg/Sachsen
Bernhard.Wagenbreth@ub.tu-freiberg.de

²⁵ http://www.dfg.de/formulare/12_102/12_102_de.pdf.