

PROF. DR. THORSTEN KOCH¹, DR. ROLF GRIEBEL²,
KONSTANZE SÖLLNER³, JÜRGEN CHRISTOF⁴, ROLAND
BERTELMANN⁵

**DeepGreen – Entwicklung eines
rechtssicheren Workflows zur
effizienten Umsetzung der
Open-Access-Komponente in den
Allianz-Lizenzen für die Wissenschaft**

¹Konrad-Zuse-Zentrum für Informationstechnik Berlin

²Bayerische Staatsbibliothek

³Friedrich-Alexander-Universität Erlangen-Nürnberg

⁴Technische Universität Berlin

⁵Helmholtz-Zentrum Potsdam - Deutsches GeoForschungsZentrum

Herausgegeben vom
Konrad-Zuse-Zentrum für Informationstechnik Berlin
Takustraße 7
D-14195 Berlin-Dahlem

Telefon: 030-84185-0
Telefax: 030-84185-125

e-mail: bibliothek@zib.de
URL: <http://www.zib.de>

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

BVB BibliotheksVerbund
Bayern

BSB Bayerische
Staatsbibliothek
Information in erster Linie

FAU FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

GFZ
Helmholtz-Zentrum
POTSDAM
Bibliothek und Informationsdienste



kobv

WISSEN IM ZENTRUM
UNIVERSITÄTSBIBLIOTHEK
Technische Universität Berlin

DeepGreen

Entwicklung eines rechtssicheren Workflows zur effizienten Umsetzung der
Open-Access-Komponente in den Allianz-Lizenzen für die Wissenschaft

Projektantrag im Bereich
“Wissenschaftliche Literaturversorgungs- und Informationssysteme”

LIS-Förderprogramm: Elektronische Publikationen

Ausschreibung: Open-Access-Transformation

Basismodul

Modul Projektspezifische Workshops

Prof. Dr. Thorsten Koch, Abteilungsleiter Wissenschaftliche Information, Leitung KOBV*
Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB),
Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV),
Takustraße 7, 14195 Berlin (Projektkoordination)

Dr. Rolf Griebel, Generaldirektor*
Bayerische Staatsbibliothek (BSB), Bibliotheksverbund Bayern (BVB),
Ludwigstraße 16, 80539 München

Konstanze Söllner, Leitung*
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Universitätsbibliothek,
Universitätsstraße 4, 91054 Erlangen

Jürgen Christof, Leitung*
Technische Universität Berlin (TU), Universitätsbibliothek,
Fasanenstr. 88 (im VOLKSWAGEN-Haus), 10623 Berlin

Roland Bertelmann, Leitung*
Helmholtz-Zentrum Potsdam – Deutsches GeoForschungsZentrum GFZ, Bibliothek und Informationsdienste,
Telegrafenberg, 14473 Potsdam

* Keine Befristung der Position

Inhalt*

1 Ausgangslage und eigene Vorarbeiten	1
1.1 Projektbezogene Publikationen	7
2 Ziele und Arbeitsprogramm	7
2.1 Voraussichtliche Gesamtdauer des Projekts	7
2.2 Ziele	8
2.3 Arbeitsprogramm und Umsetzung	9
2.4 Maßnahmen zur Erfüllung der Förderbedingungen / Umgang mit Projektergebnissen	22
3 Literaturverzeichnis	22

Alle übrigen Teile des Antrages (Finanzierung des Vorhabens, beteiligte Personen, etc) sind nicht Teil dieser Veröffentlichung.

* Gliederungspunkte, die für den Antrag nicht von Bedeutung waren, wurden nicht mit Inhalt gefüllt. Daher ist die Gliederung kürzer als in der Muster-Vorhabensbeschreibung vorgegeben.

1 Ausgangslage und eigene Vorarbeiten

Open Access: Bedeutung des Grünen Wegs

Die Diskussion um die verschiedenen Wege von Open Access, hin zu Gold oder Grün, hat sich in den letzten Jahren sehr dynamisch entwickelt. Obwohl im internationalen Kontext Gold intensiv diskutiert wird, scheint der Weg der Förderorganisationen aktuell Grün weiter zu stärken: Grants des Europäischen Forschungsrats (ERC), Horizon 2020 und die sieben spezifischen Programme des 7. EU-Forschungsrahmenprogramms (FRP) verpflichten zum grünen Open-Access-Archiving – dem zeitgleichen oder nachträglichen Veröffentlichen von Verlagspublikationen auf der Homepage des Autors oder auf einem Open-Access-Dokumentenserver (Repositorium).

Großbritannien agiert wegweisend, indem Verpflichtungen ausgesprochen werden: Ab dem 1. April 2016 zählen im Rahmen des Research Excellence Framework (REF) für das Higher Education Funding Council for England (HEFCE) bei der Bewertung von Hochschulen und deren Impact Faktor Veröffentlichungen nur noch dann, wenn sie binnen dreier Monate in ein Open-Access-Repositorium eingespeist werden.¹ Diese bindende Voraussetzung wird dazu führen, dass wissenschaftliche Publikationen aktiv und zeitnah ins Repositorium eingestellt werden – idealerweise direkt durch den Verlag.

Auch in Deutschland sind wichtige Akzente gesetzt worden: Eine Reihe von Bibliotheken haben mit finanzieller Unterstützung der Deutschen Forschungsgemeinschaft (DFG) und im Rahmen der Schwerpunktinitiative "Digitale Information" der Allianz der deutschen Wissenschaftsorganisationen seit 2011 sogenannte Allianz-Lizenzen mit Verlagen verhandelt, in denen weitreichende Rechte hinsichtlich der Open-Access-Archivierung für AutorInnen verankert sind. Auch in den bis 2010 von der DFG geförderten Nationallizenzen wurden in den späteren Jahren diese Rechte teilweise vereinbart. AutorInnen zugriffsberechtigter Einrichtungen können Artikel, die in den im Rahmen der DFG-geförderten Verträge lizenzierten Zeitschriften erschienen sind, ohne oder mit nur kurzer Embargofrist in geeigneten Repositorien ihrer Wahl frei zugänglich machen. Zumeist dürfen dafür die von den Verlagen erstellten PDF-Dateien verwendet werden. Die Verhandlung der Open-Access-Komponente stellt nach den Grundsätzen der DFG für den Erwerb überregionaler Lizenzen eine zwingende Fördervoraussetzung dar.²

Geringe Inanspruchnahme der verhandelten Open-Access-Komponente aus den Allianz-Lizenzen in Deutschland

Die Erfahrung der seit 2011 getätigten Allianz-Abschlüsse³ zeigt allerdings, dass der Kreis berechtigter AutorInnen nur sehr zögerlich Gebrauch von seinen Open-Access-Rechten macht. 2013 hat eine Umfrage der Bayerischen Staatsbibliothek (BSB) unter den teilnehmenden Einrichtungen ergeben, dass die Initiative zur Publikation fast ausschließlich von den an den Allianz-Lizenzverträgen beteiligten Bibliotheken ausging.⁴ Aber auch hier ist die aktive Umsetzung der verhandelten Open-Access-Komponente eher der Ausnahmefall.

¹ Vgl. <http://www.hefce.ac.uk/pubs/year/2014/201407/#d.en.86771> (Zugriff am 28.10.2014).

² Vgl. dazu http://www.dfg.de/formulare/12_181/12_181_de.pdf (Zugriff am 28.10.2014).

³ In den letzten Jahren der Förderung von Nationallizenzen, die 2010 auslief, wurden die Rechte zur Zweitverwertung vertraglich auch für abgeschlossene Zeitschriftenarchive gesichert.

⁴ 166 Teilnehmer, 44 Rückmeldungen, 16 davon aktiv gewordene Einrichtungen im oben beschriebenen Sinne. Insgesamt waren zum Zeitpunkt der Umfrage rund 600 Beiträge in Repositorien abgelegt worden. Neben dem Fehlen entsprechender Repositorien war insbesondere das mangelnde Interesse der AutorInnen als Grund für die geringe Verwertung genannt worden (Informationen der Bibliotheken über die Open-Access-Komponente verhalten bei den WissenschaftlerInnen ungehört).

Gründe dafür sind, dass erstens für die Bibliotheken als Betreiber von Repositorien die verhandelten Rechte zwar dokumentiert⁵, aber relativ unübersichtlich sind, da sich diese von Vertrag zu Vertrag unterscheiden; zweitens nur wenige Institutionen ihren AutorInnen Betreuung bzw. technische Unterstützung bei der effektiven Umsetzung der Open-Access-Zweitveröffentlichung anbieten können, so dass die WissenschaftlerInnen den Mehraufwand scheuen; und drittens, es für Repositorienbetreiber schwierig ist, aus den Verlagsdaten selbst eine Selektion nach AutorInnen aus einer bestimmten Institution (Affiliation) zu ermitteln, da die Angaben in den Metadaten oft nicht normiert sind und kein automatisierter Download von Metadaten oder Volltext möglich ist. Eine Übernahme der Publikationen in Repositorien erfolgt deshalb – wenn überhaupt – nur einzeln und manuell.

Nach wie vor liegt also ein großer Schatz wissenschaftlicher Literatur allein bei den Verlagen. In Zahlen bedeutet das: 18 an Allianz- und Nationallizenzen beteiligte Verlagshäuser publizieren insgesamt 2.194 Zeitschriften, deren AutorInnen – sofern sie einer der zeichnenden Institutionen angehören – zur grünen Open-Access-Veröffentlichung berechtigt wären, diese Möglichkeit aber fast nie nutzen.⁶

Beispielhafte Umsetzung der Open-Access-Komponente aus Allianz-Lizenzen beim Projektpartner FAU

Die Universitätsbibliothek der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) wertet seit 2011 sieben Allianz-Verträge⁷ regelmäßig auf Artikel von Universitätsangehörigen aus. Nach Ablauf eventueller Embargofristen werden die WissenschaftlerInnen zur Veröffentlichung dieser Artikel in das institutionelle Repository OPUS FAU aufgefordert.⁸ Die Anfragen werden von AutorInnenseite regelmäßig positiv beantwortet.

Quantitativ konnten auf diese Weise im Schnitt 116 Artikel pro Jahr im OPUS FAU im Grünen Weg publiziert werden. Der investierte Aufwand hierfür ist jedoch sehr hoch: Die Suche nach AutorInnen der eigenen Einrichtung ist nicht bei allen Verlagen möglich, da entsprechende Suchfelder fehlen. Die alternative, monatliche Sichtung aller Verlagserzeugnisse auf FAU-AutorInnen ist mit den vorhandenen Personalkapazitäten selten durchführbar, wodurch in der Folge nicht alle Artikel mit FAU-Beteiligung erfasst werden können. Eine bibliometrische Analyse über die getroffene Verlagsauswahl ergab für das Jahr 2013 eine weitaus höhere tatsächliche Zahl der Artikel mit FAU-Beteiligung: insgesamt 291, von denen nur etwa 40% (116 \cong 39,86% von 291) der publizierten Artikel schließlich in OPUS FAU erfasst wurden.

Den wissenschaftlichen Schatz heben mit DeepGreen

Mit DeepGreen verfolgen die Antragssteller das Ziel, einen Großteil jener Aufsätze, die unter den speziell im DFG-geförderten Kontext verhandelten Bedingungen grün online gehen dürften, auch tatsächlich online abrufbar zu machen. DeepGreen agiert dabei als Intermediär zwischen abliefernden Verlagen und abholenden institutionellen und fachlichen Open-Access-Repositorien. Dazu soll mit den Verlagen und Bibliotheken – als Betreiber der Repositorien und damit Vertreter für die berechtigten AutorInnen – ein Workflow entwickelt und ein automatisiertes Ablieverfahren geprüft werden. Als eines der technischen Elemente ist ein Repository angedacht, das in der

⁵ Vgl. dazu <http://www.nationallizenzen.de/open-access/open-access-rechte.xls/view> (Zugriff am 28.10.2014).

⁶ Daneben steht die Zahl jener Institutionen, die nach dem Prinzip der so genannten *Moving Wall* nach Ablauf jeweils vereinbarter Embargofristen das kostenfreie Archiv nutzen und ab diesem Zeitpunkt ebenfalls mit ihren AutorInnen das Recht haben, Open Access zu publizieren. Hier sei nur auf ein einzelnes Beispiel, Sage Publications, verwiesen: 135 Institutionen haben das kostenpflichtige Allianz-Lizenzpaket gezeichnet, 414 aber nutzen das Archiv, das über die *Moving Wall* laufend erweitert wird, als Nationallizenz und erhalten damit nach Ablauf des Embargos ebenfalls die Möglichkeit, Artikel in ihr Repository zu überführen.

⁷ AIP Journals, BMJ Journals, Cambridge Journals, IOPscience, Oxford Journals, Karger, Sage Publications.

⁸ Die entsprechenden Sammlungen („Collections“) sind unter <http://opus4.kobv.de/opus4-fau/solrsearch/index/search/searchtype/collection/id/16213> (Zugriff am 28.10.2014) zu finden.

beantragten Projektphase als reines Dark Archive agieren wird und als "Verteilerzentrum" für die Dokumente und Metadaten dient. In einem zweiten Schritt wird geprüft, ob sich DeepGreen perspektivisch hin zu einem zentralen, fachübergreifenden, nationalen Repositorium entwickeln kann, mit nach außen sichtbarem Content und Schaufensterwirkung.

Das Projekt DeepGreen stützt sich auf international offene, standardisierte Schnittstellen und Metadatenformate (z.B. SWORD <http://swordapp.org/>) und fördert deren Anwendung in Deutschland. Mit DeepGreen wird die in den Allianz-Lizenzverträgen vereinbarte Open-Access-Komponente vorbildhaft und vertrauensbildend umgesetzt, WissenschaftlerInnen entsteht keine Mehrarbeit und Bibliotheken werden zukünftig entlastet. Dafür muss als wichtigste Voraussetzung eine Vertrauensbasis mit den beteiligten Verlagen geschaffen werden. Eingeworben werden soll die aktive Mitarbeit der Verlage, zu der sie vertraglich keineswegs verpflichtet sind. Damit verbunden ist die Erwartung, dass die Akzeptanz des grünen Open-Access-Publikationswegs steigt, wenn abgestimmte Workflows existieren. Perspektivisch soll mit DeepGreen nicht nur eine Infrastruktur für die Ermittlung, den Abzug und die Verteilung der relevanten Daten an die berechtigten Einrichtungen bzw. deren Repositorien entstehen, sondern potentiell auch eine zentrale Zugriffsplattform.

DeepGreen als Best Practice

Ziel ist es, einen rechtssicheren, vertrauensvollen Workflow zu erarbeiten, der für alle Beteiligten (Verlage, Bibliotheken als Betreiber von fachlichen und institutionellen Repositorien und nicht zuletzt die wissenschaftlichen AutorInnen) akzeptabel ist bzw. eine Arbeitserleichterung darstellt und gleichzeitig die Anzahl der Open-Access-Beiträge in den Repositorien signifikant erhöht. Die Nutzung dieses Workflows und des DeepGreen-Repositoriums könnte zur Gewährleistung einer langfristigen Nachnutzung auch in die DFG-Vergaberichtlinien einfließen und ebenso für konsortiale und lokal vereinbarte Open-Access-Komponenten nutzbar sein.

Bekannte Projekte und konzeptionelle Vorbilder

Das Vorhaben DeepGreen nimmt sich konzeptionell folgende europäische Projekte zum Vorbild:

EUROPE PubMed Central (Europe PMC, <http://europepmc.org/>) und **PubMed Central International** (PMCI, <http://www.ncbi.nlm.nih.gov/pmc/about/pmci/>) stellen für die Medizin, Biologie und angrenzende Lebenswissenschaften international Open-Access-Publikationen zur Verfügung. Die Ablieferung durch die Verlage erfordert eine Validierung nach einem beschriebenen Metadatenformat, sie erfolgt über FTP und setzt damit bewusst niedrige technische Hürden.⁹ Für die Weiterverwertung und den Download der Open-Access-Inhalte existieren gerade bei dem technisch avancierteren Europe PubMed Central ein ganzer Strauß unterschiedlicher Möglichkeiten (von SOAP über OAI, FTP oder API). Die Ablieferungswege seitens der Verlage sind vertraglich geregelt und eingeübt. Funktional können diese Infrastrukturen für DeepGreen als vorbildhaft gelten.

Eng verwandt mit DeepGreen ist das europäische **PEER Project**, 'Publishing and the Ecology of European Research', in dem von 2009 bis 2012 Vertreter von STM-Verlagen und Förderer erstmals eine große Infrastruktur für den Grünen Weg erprobt haben. Das Projekt wurde im Frühjahr 2012 beendet (<http://www.peerproject.eu/>). Dabei ist auch deutlich geworden, dass die zu leistenden Aufwände enorm ("Delivery of a working Green Open Access environment, even for an experiment, is highly complex and challenging", Zitat aus Abschlusspräsentation¹⁰), und die Zusammenarbeit aller Beteiligten unabdingbare Voraussetzung sind. Die konzeptionelle Basis, die im PEER Project gelegt wurde (Spezifikation der Workflows, Monitoring, Heterogenität der Daten, Schnittstellen), soll für DeepGreen nachgenutzt werden. Im Unterschied zum PEER Project, bei

⁹ <http://www.ncbi.nlm.nih.gov/pmc/pub/pubinfo/#techreqs> (Zugriff am 28.10.2014).

¹⁰ http://www.peerproject.eu/fileadmin/media/presentations/PEER-Executive_Partners_Statements_29_May_2012.pdf (Zugriff am 28.10.2014).

dem es sich um eine komplette Eigenentwicklung handelte, soll sich DeepGreen auf eine (noch auszuwählende) existierende Software stützen.

Im Rahmen von **SCOAP3** (<http://scoap3.org/>), 'Sponsoring Consortium for Open Access Publishing in Particle Physics', wurden zentrale Fachzeitschriften der Hochenergiephysik in Open Access transformiert. Alle über das internationale Konsortium finanzierten Artikel werden von den Verlagen nach einem definierten Verfahren¹¹ im SCOAP3 Repository (<http://scoap3.org/scoap3-repository>) zugänglich gemacht. Auch hier sind wie bei Europe PMC verschiedene technische Ablieferwege vorgesehen, gleichzeitig wird ein Metadatenformat vorgegeben. Die verwendeten Standards und Erfahrungen können als Grundlage für die Konzeption von DeepGreen dienen.

Konkrete Vorarbeiten der Projektbeteiligten

Das nationale Projektkonsortium besteht aus zwei Bibliotheksverbänden, zwei Universitäten, der Bayerischen Staatsbibliothek und einer außeruniversitären Forschungseinrichtung. Die beiden Verbände – der Kooperative Bibliotheksverbund Berlin-Brandenburg (KOBV) und der Bibliotheksverbund Bayern (BVB) – engagieren sich im Projekt mit ihrer Entwicklungspartnerschaft.

KOBV. Der Kooperative Bibliotheksverbund Berlin-Brandenburg (KOBV) wirkt für die Bibliotheken der Metropolregion Berlin-Brandenburg als Dienstleistungs- und Entwicklungszentrum. Grundlegender methodischer Ansatz des Verbundes ist dabei die enge Verzahnung und gemeinsame Durchführung von Betriebs- und Entwicklungsaufgaben in der KOBV-Zentrale. Service und Entwicklung werden zum gegenseitigen Nutzen zukunftsorientiert miteinander verbunden. Zu diesem Zweck ist die KOBV-Zentrale mit einem eigenen Entwicklungssektor ausgestattet, in dem kontinuierlich Forschungs- und Entwicklungsprojekte durchgeführt werden.

In einem dieser Projekte wurde mit dem „KOBV-Volltextserver“ prototypisch ein Angebot geschaffen, das die verlagsunabhängige Sicherung und die langfristige Verfügbarkeit von Zeitschriftenvolltexten gewährleistet. Dazu wurde mit der zeitgleich entwickelten Weboberfläche „ReMaS“ ein komfortables und transparentes IP-basiertes Rechtemanagementsystem entwickelt. Der Volltextserver bietet den institutsbeschränkten (IP-regulierten) Download von 1,7 Mio. lizenzierten Zeitschriftenartikeln, sowie den freien Zugang zu 35.000 Open-Access-Publikationen an. Im Rahmen des Projektes wurden vertrauensvolle Beziehungen zu Verlagen geschaffen, gleichzeitig konnte die Verarbeitung heterogener Datenlieferungen (Metadaten und Volltexte) von mehreren Verlagen getestet werden.

Zur Unterstützung der Bibliotheken bei ihren Open-Access-Aktivitäten entwickelt die KOBV-Zentrale seit Sommer 2010 federführend die Open-Source-Software OPUS 4. Anforderungen werden in einem schnellen, agilen Entwicklungszyklus umgesetzt. OPUS 4 ist mittlerweile in 50 Einrichtungen im Einsatz, 26 davon werden vom OPUS-Hostingteam in der KOBV-Zentrale betreut (Stand: 10/2014).

Rolle innerhalb des Projekts: Der KOBV übernimmt in diesem Vorhaben die Gesamtprojektkoordination und verantwortet die technische Spezifikation und prototypische Implementierung des zentralen technischen Bausteins, der Datendrehzscheibe. Die Verbundzentrale wirkt an der Erarbeitung des Metadatenschemas und der Konzeption des Gesamtworkflows mit, realisiert notwendige Erweiterungen in der OPUS-Software als institutionelles Repositorium und unterstützt bei der fachlichen Vernetzung (Öffentlichkeitsarbeit).

Bayerische Staatsbibliothek/ Bibliotheksverbund Bayern. Die Bayerische Staatsbibliothek (BSB) kann auf langjährige Erfahrungen bei der Verhandlung, Verwaltung und Bereitstellung auch großvolumiger Lizenzabschlüsse im regionalen, überregionalen und internationalen Kontext

¹¹ Siehe: http://scoap3.org/files/Technical_Specification.pdf (Zugriff am 28.10.2014).

zurückgreifen. Hier ist die Geschäftsstelle des Bayern-Konsortiums angesiedelt, in der zahlreiche Lizenzverträge für die wissenschaftlichen Bibliotheken des Freistaats verhandelt und administriert werden. Ebenso werden Konsortiallizenzen mit überregionalem Teilnehmerkreis abgeschlossen. Aktuell werden 82 Lizenzabschlüsse mit 1.065 Teilnehmerlizenzen betreut.

Die Bayerische Staatsbibliothek ist Gründungsmitglied der Arbeitsgemeinschaft Deutscher, Österreichischer und Schweizer Konsortien (GASCO) und hat dort den stellvertretenden Vorsitz inne. Sie ist darüber hinaus von Beginn an verhandlungsführend und konzeptionell an der Ausgestaltung der DFG-geförderten National- und Allianz-Lizenzen beteiligt. Betreut werden über 40 National-Lizenzabschlüsse und 13 Allianz-Lizenzen. Auf Einladung der DFG ist die Bayerische Staatsbibliothek in der Allianz-AG Lizenzen vertreten. Regelmäßig wirkt sie bei der Entwicklung von Lizenzierungsstandards mit, z.B. in einer Unter-AG zur Erarbeitung einer mittlerweile vorliegenden Musterlizenz. Aus dem Allianz-Kontext erwachsen ist das gemeinsam mit der Zentralbibliothek des Forschungszentrums Jülich (FZJ) durchgeführte DFG-Projekt zur „Entwicklung eines Klassifikationsschemas (Banding) für die Informationsbeschaffung für wissenschaftliche Einrichtungen in Deutschland“. Die BSB wirkt darüber hinaus aktiv in der Programmgestaltung der jährlichen Tagungen der ‘International Coalition of Library Consortia’ (ICOLC) mit.

Am 2008 gegründeten “Zentrum für Elektronisches Publizieren” hat die BSB in den vergangenen Jahren für den vorliegenden Antrag Erfahrungen im Bereich des Open-Access-Publizierens wissenschaftlicher Inhalte diverser Art gesammelt (Zeitschriften, Rezensionen, Primärdaten, Blogs). Neben technischer und organisatorischer Expertise können hier insbesondere Erfahrungen eingebracht werden, die die Interessenberücksichtigung und ggf. Einbindung der beteiligten AutorInnen sowie die juristischen Implikationen der Onlinestellung betreffen.

Mit der Verbundzentrale des Bibliotheksverbands Bayern (BVB) ist die Bayerische Staatsbibliothek zudem für die Versorgung aller wissenschaftlichen Bibliotheken Bayerns mit zukunftsfähigen Informationsinfrastrukturen verantwortlich. Die BVB-Verbundzentrale betreut das Verbundportal Gateway Bayern mit über 85 Millionen Datensätzen. Während des Aufbaus konnten hinreichend Erfahrungen bei der Integration heterogener Metadaten in einen einheitlichen Suchraum gesammelt werden. Mit der seit 2007 bestehenden gemeinsamen Entwicklungspartnerschaft von BVB und KOBV wird die Entwicklung leistungsfähiger Bibliotheksanwendungen vorangetrieben.

Rolle innerhalb des Projekts: Die BSB übernimmt in diesem Projekt die Kommunikation mit den Verlagen und vernetzt das Projekt in nationalen und internationalen Konsortialzusammenhängen (GASCO/ ICOLC). Das “Zentrum für Elektronisches Publizieren” der BSB beteiligt sich an der Workflowentwicklung, die auch die Erprobung von Dateneinspeisungen in die von der BSB betriebenen Repositorien umfasst. Die Verbundzentrale des BVB unterstützt die technische Spezifikation und Implementierung.

FAU. Die Universitätsbibliothek Erlangen-Nürnberg ist als zentraler Informationsdienstleister ein wichtiger Initiator für Open-Access-Aktivitäten an der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). Sie ist seit 2010 verantwortliche Ansprechpartnerin für den DFG-geförderten Fond “Open Access Publizieren”, hat die Open-Access-Infrastruktur vor Ort auf- und ausgebaut (OPUS FAU, FAU University Press) und hochschulpolitische Entscheidungen angeregt und unterstützt, wie die Einführung einer Open-Access-Policy sowie die Unterzeichnung der Berliner Erklärung. Jüngster Erfolg: In einem Neubeschluss der Universitätsleitung werden alle neuen WissenschaftlerInnen bei der Berufung bzw. Einstellung der FAU darauf aufmerksam gemacht, den Grünen Weg zu beschreiten.

Zudem trat die FAU als Institutional Member der Confederation of Open Access Repositories (COAR) bei. Die im Rahmen von Allianz-Lizenzen verhandelten Open-Access-Rechte werden

durch die Universitätsbibliothek proaktiv in Anspruch genommen (s.o.). Das Repositorium OPUS FAU wird durch den Bibliotheksverbund KOBV gehostet, mit dem sich in den letzten Jahren eine intensive Zusammenarbeit etablierte. Durch Gewährung eines Systemzugangs konnte das Update von OPUS 3 auf OPUS 4 in kürzester Zeit kooperativ mit dem KOBV durchgeführt werden, d. h. die Mitarbeiter der Universitätsbibliothek Erlangen-Nürnberg haben im begrenzten Rahmen entsprechende Systemerfahrungen; OPUS FAU war anschließend Pilotinstanz für die Tests zur XMetaDissPlus-Schnittstelle.

Rolle innerhalb des Projekts: Unterstützung bei der Anforderungserhebungen aus Bibliotheks- und Repositoriensicht, Pilotpartner als institutioneller Datenabnehmer (OPUS FAU), Erhebung und Auswertung von Kennzahlen und Öffentlichkeitsarbeit. Die FAU wird die Koordination bei der Entwicklung eines Metadatenschemas übernehmen.

TU Berlin. Die Universitätsbibliothek der Technischen Universität Berlin (TU Berlin) ist die größte technisch-naturwissenschaftliche Bibliothek der Region Berlin-Brandenburg. Als zentrale Informationseinrichtung ist sie zuständig für die Beschaffung und Bereitstellung von Medien und Informationen in gedruckter und elektronischer Form. Darüber hinaus ist es ihre Aufgabe, die Mitglieder der TU Berlin in allen Fragen des Publizierens und Archivierens zu unterstützen. Sie ist Initiator wesentlicher Open-Access-Aktivitäten und stellt den WissenschaftlerInnen der TU Berlin zahlreiche Serviceleistungen rund um das Thema Open-Access-Publizieren zur Verfügung. Zur Open-Access-Infrastruktur der Universitätsbibliothek der TU Berlin zählen das Digitale Repositorium, das Forschungsdatenrepositorium DepositOnce, der Universitätsverlag der TU Berlin, der Zeitschriftenserver und ein umfangreiches Beratungsangebot.

Das Digitale Repositorium der TU Berlin umfasst derzeit über 4.300 Dokumente (Stand: 9/2014), es läuft unter OPUS 4 und wird derzeit vom KOBV gehostet. Im Jahr 2015 wird es zu DepositOnce migrieren. DepositOnce, das Repositorium für Forschungsdaten und Publikationen der TU Berlin, ist seit August 2014 in Routinebetrieb. Es dient den WissenschaftlerInnen als Serviceplattform für die langfristige Aufbewahrung und Verfügbarkeit ihrer Forschungsergebnisse. DepositOnce nutzt die Open-Source-Repository-Software DSpace.

Mit dem Universitätsverlag der TU Berlin steht den WissenschaftlerInnen der Universität seit über zehn Jahren ein eigener Open-Access-Verlag zur Verfügung. Er hat das Ziel, Forschungsergebnisse der TU Berlin unter einheitlichem Layout professionell aufzubereiten und über das Internet sowie bei Bedarf über die traditionellen Vertriebswege weltweit zu verbreiten. Der Verlag gibt jährlich ca. 70 Neuerscheinungen in gedruckter und/oder elektronischer Form heraus.

Die Universitätsbibliothek unterhält einen Zeitschriftenserver zum Hosting von begutachteten Open-Access-Zeitschriften auf der Basis der Publikationsplattform „Open Journal Systems“. Eine erste Zeitschrift ist bereits online. Die Universitätsbibliothek bietet den AutorInnen umfangreiche Beratung beim Open-Access-Publizieren an, insbesondere unterstützt sie nachdrücklich bei der Klärung ihrer Zweitveröffentlichungsrechte.

Rolle innerhalb des Projekts: Die UB der TU Berlin wirkt mit an der Erarbeitung des Metadatenschemas und an der technischen Spezifikation. Sie ist als institutioneller Datenabnehmer für die Anbindung der Repositoriumssoftware DSpace verantwortlich. Bei der Kennzahlenerhebung und Auswertung der Projektergebnisse übernimmt sie die Koordination.

GFZ Potsdam. Als nationales Forschungszentrum für die Geowissenschaften beschäftigt sich das Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum (GFZ) mit dem System Erde. Dabei wird es durch die Abteilung Bibliothek und Informationsdienste (LIS) mit dienstleistungsorientierten, innovativen digitalen Diensten der Informationsversorgung und -verarbeitung unterstützt.

In ihrer Rolle als Bibliothek für den gesamten Wissenschaftspark Albert Einstein betreibt sie diverse Open-Access-Dienste, z.B. Repositorien und einen Open-Access-Verlag. Von 2010 bis 2013 wurden so durchschnittlich 37% aller peer-reviewten Artikel des GFZ offen zugänglich gemacht.

Als Helmholtz-Zentrum unterstützt das GFZ den offenen Zugang zu wissenschaftlichem Wissen entsprechend der Berliner Erklärung. Seit 2005 ist das Helmholtz Open Science Koordinationsbüro am GFZ angesiedelt. Diese Einrichtung koordiniert die Umsetzung von Open Science in der Helmholtz-Gemeinschaft. Die am GFZ involvierten MitarbeiterInnen sind dabei in nationale und internationale Open-Access-Initiativen eingebunden. Dazu gehören neben relevanten Open-Access-Arbeitsgruppen der Schwerpunktinitiative "Digitale Information" der Allianz der deutschen Wissenschaftsorganisationen auch die Arbeitsgruppe Elektronisches Publizieren der Deutschen Initiative für Netzwerkinformation (DINI).

Weiter ist das GFZ in diverse Entwicklungsprojekte der Informationsinfrastruktur eingebunden, z.B. ALBERT – All Library Books, Journals and Electronic Records Telegrafenberg (gemeinsam entwickelt mit dem KOBV) und re3data.org – Registry of Research Data Repositories. Darüber hinaus fördert die Bibliothek seit vielen Jahren die Entwicklung von Open Access in den Geowissenschaften, u. a. im Kontext des Arbeitskreises der Geo- und Umweltbibliotheken und der Initiative Geo.X.

Rolle innerhalb des Projekts: Das GFZ stellt den Dialog zur wissenschaftlichen Öffentlichkeit und zur Informationsinfrastruktur sicher und bringt seine Erfahrung im Bereich des wissenschaftlichen Publikationsmanagements ein (Anforderungsanalyse). Als Betreiber von ALBERT leistet das GFZ einen Beitrag zur Entwicklung des Datenschemas und erprobt die Implementierung des Workflows für die Repository-Software eSciDoc/PubMan. Das GFZ übernimmt die Koordination der Testläufe und der Implementierung von Erweiterungen in den institutionellen Repositorien.

1.1 Projektbezogene Publikationen

Beucke, D., David, S., Scholze, F., Pampel, H., Voigt, M., Zielke, D., Müller, U., Meinecke, I., Hartmann, T., Mruck, K., Vierkant, P., Steinke, T., Summann, F. (2013). DINI-Zertifikat für Open-Access-Repositorien und -Publikationsdienste 2013. DINI Schriften 3-de. Version 4.0. [urn:nbn:de:kobv:11-100217162](https://nbn-resolving.org/urn:nbn:de:kobv:11-100217162).

Kuberek, M. (2013). Die Forschungsdaten-Infrastruktur der TU Berlin. Bibliotheksdienst, 47(11), S. 833-846. <http://dx.doi.org/10.1515/bd-2013-0096>.

Pampel, H., Bertelmann, R. (2009). Open Access – neue Wege in der Wissenschaftskommunikation. GAIA – Ecological Perspectives for Science and Society, 18(4), S. 337-339. <http://gfzpublic.gfz-potsdam.de/pubman/item/escidoc:239762>

Pampel, H., Bertelmann, R., Hübner, A. (2009): Aktionsfeld Open Access. Wissenschaftsmanagement: Zeitschrift für Innovation, 15(2), S. 11-17. <http://gfzpublic.gfz-potsdam.de/pubman/item/escidoc:238794>

Schäffler, H. (2014): Elektronische Medien in der überregionalen Literaturversorgung: Nationallizenzen, Allianz-Lizenzen, FID-Lizenzen. - In: Göttker, S. / Wein, F.: Neue Formen der Erwerbung, De Gruyter, S. 204-222.

Sitek, D., Bertelmann, R. (2014): Open Access: a state of the art. - In: Sönke, B., Friesike, S. (Eds.), Opening Science: the evolving guide on how the internet is changing research, collaboration and scholarly publishing, Springer, S. 139-154. DOI: http://doi.org/10.1007/978-3-319-00026-8_9.

2 Ziele und Arbeitsprogramm

2.1 Voraussichtliche Gesamtdauer des Projekts

Beantragt wird zunächst eine Dauer von **zwei Jahren** für die Entwicklung eines prototypischen Workflows. Bei erfolgreichem Abschluss wird ein Antrag auf Verlängerung um **zwei Jahre** gestellt. Ziel der zweiten Projektphase wäre es, aufbauend auf den Erkenntnissen den DeepGreen-Workflow auszubauen und Wege in die Nachhaltigkeit zu ebneten.

2.2 Ziele

Konkrete Ergebnisse des Vorhabens für die Optimierung der Informationsinfrastruktur

Erarbeitet wird ein zwischen Verlagen und Bibliotheken abgestimmter Workflow zur nahezu vollständig automatisierten Ablieferung von Open-Access-Publikationen. Als technisches Hilfsmittel dient ein intermediäres Repositorium. Als Pilot ist dieses Repositorium zunächst ein Dark Archive, das im Hintergrund als Datendrehscheibe agiert. Prototypisch wird mit Verlagen kooperiert, mit denen bereits umfangreiche Open-Access-Rechte im Rahmen von Allianz-Lizenzen verhandelt wurden. Es wird ein Metadatenschema abgestimmt und ein Verfahren erarbeitet, das Bibliotheken berechtigt, umfassend für die AutorInnen der eigenen Einrichtung tätig zu werden. Im Ergebnis werden die wissenschaftlichen Ergebnisse umfassender als bislang online frei verfügbar sein, ohne dass von den AutorInnen zusätzliche Arbeit erforderlich ist.

Innovatives Element des Vorhabens

Neu ist der Grad der Automatisierung. Angedacht ist ein automatisierter Datenablieferungsprozess von Verlagen an ein Dark Archive (perspektivisch: nationales Repositorium) und ausgehend von dieser intermediären Drehscheibe an Repositorien. Es wird ein technisch offenes System konzipiert, das auf international üblichen, offenen Schnittstellen beruht und sich funktional an führenden, internationalen Vorbildern wie Europe PMC und SCOAP3 orientiert.

Potential für die Transformation der Wissenschaftskommunikation

DeepGreen soll der erste Schritt zu einem sichtbaren, nationalen, fachübergreifenden Open-Access-Repositoriums sein. WissenschaftlerInnen profitieren durch die starke Vereinfachung des Grünen Wegs: Ihre Publikationen werden ohne zusätzlichen Aufwand Open Access gestellt und erreichen eine weltweite Leserschaft. Langfristig könnte die neue Infrastruktur auch über die verhandelten Open-Access-Rechte in den hier wesentlichen Allianz-Lizenzverträgen hinaus an Bedeutung gewinnen: Zukünftig wären besonders bei Publikationen aus Drittmittelprojekten entsprechende Zusätze zu Autorenverträgen denkbar, um Compliance mit den Open-Access-Policies der Drittmittelgeber zu garantieren. Neben autorensseitigen Verträgen kann auch eine Vielzahl von bibliotheksseitigen Verhandlungen, auch über Konsortial-Lizenzen, die Option der Nachnutzung des Dienstes für die Automatisierung des Grünen Wegs vorsehen.

Schließlich ist es möglich, für Bibliotheken aus einer Verbundregion die relevanten Metadaten en bloc einer Verbundzentrale zur Verfügung zu stellen oder auf Updates von Metadaten und Volltext hinzuweisen. Ausgehend von einem freien Angebot der Metadaten ist ggf. auch die technische Unterstützung des Exports von Metadaten aus DeepGreen in Universitätsbibliographien und Forschungsinformationssysteme – gefiltert nach der Affiliation von Autoren – von Interesse.

Unterschiede zu vergleichbaren Lösungsansätzen

Ein zentrales, fachübergreifendes Repositorium für wissenschaftliche Verlagsartikel, das Metadaten und auch Volltexte enthält, diese automatisiert verarbeitet und weiterverteilt, existiert in Deutschland bislang nicht. Als Basisdienstleistung werden von dem Gemeinsamen Bibliotheksverbund in Göttingen (GBV) ausschließlich die Verlags-Metadaten in gängigen bibliographischen Formaten (MARC21, MAB) zum Download bereitgestellt und IP-Zugänge verwaltet (<https://www.nationallizenzen.de/>). Volltexte im Open Access sind hier nicht enthalten. Erste Ablieverfahren werden gerade für die Pflichtabgaben an die Deutsche Nationalbibliothek (DNB) erprobt. Etablierte, vertraglich bindende Ablieverfahren von Verlagen sind allerdings bislang nur

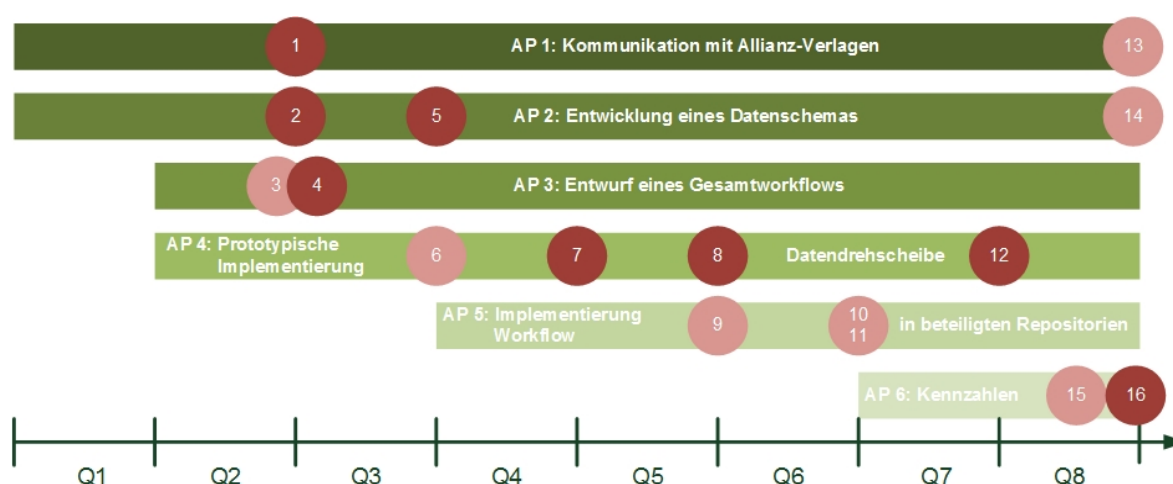
außerhalb Deutschlands an fachliche Repositorien wie Europe PMC oder SCOAP3 bekannt. Darüber hinaus liefern Open-Access-Verlage via <http://swordapp.org/> an Repositorien.¹²

Risikoabschätzung und Risikomanagement

Ein Risiko des Projekts liegt in der Bereitschaft der Verlage, sich mit automatisierten Ablieferungsverfahren an der geplanten neuen Infrastruktur zu beteiligen. Hier ist eine prinzipielle Skepsis denkbar ebenso wie reale, technisch-organisatorische Schwierigkeiten. Diesem Risiko begegnet das Vorhaben mit der Entscheidung für ein anfängliches „Dark Archive“, das in keiner „sichtbaren“ Konkurrenz zu den Verlagsangeboten steht. Unter dieser Voraussetzung konnten mit Sage Publications und Karger bereits zwei Verlage für eine Zusammenarbeit im Projekt DeepGreen gewonnen werden. Ein weiterer Verlag (BMJ) hat sein grundsätzliches Interesse bekundet. Hinsichtlich der Metadaten, ihrer Konsistenz und Qualität muss seitens der Bibliotheken pragmatisch gedacht und mit Abstrichen gerechnet werden. Auch gilt es, das Verhältnis zwischen Implementierungsaufwand eines intermediären Repositoriums und seinem realen Nutzen kritisch zu bewerten. Dieser Frage ist in diesem Vorhaben ein eigenes Arbeitspaket (AP 6: Evaluierung) gewidmet.

2.3 Arbeitsprogramm und Umsetzung

Übersicht über Zeitplanung und Meilensteine



Nr.	Meilenstein (fett= externer Meilenstein, normal = interner Meilenstein)	Arbeitspaket	Quartal
1.	Testdatenlieferungen von mindestens 2 Verlagen	AP 1	Q2
2.	Durchführung Experten-Workshop	AP 2	Q2
3.	Erarbeitung eines Workflows (Version 1.0)	AP 3	Q2
4.	Veröffentlichung eines Datenflussdiagramms	AP 3	Q2
5.	Veröffentlichung von Metadatenschema Version 1.0	AP 2	Q3
6.	Festlegung technischer Werkzeuge und Schnittstellen für DeepGreen	AP 4	Q3
7.	Veröffentlichung Prototyp Version 1.0: Datenübernahme	AP 4	Q4
8.	Veröffentlichung Prototyp Version 2.0: Datenabgabe	AP 4	Q5
9.	Test: Verlagsdatenübernahme mit OPUS 4 (FAU)	AP 5	Q5
10.	Test: Verlagsdatenübernahme mit DSpace (TU Berlin)	AP 5	Q6
11.	Test: Verlagsdatenübernahme mit eSciDoc/PubMan (GFZ)	AP 5	Q6
12.	Veröffentlichung Prototyp Version 3.0 (Bugfixes)	AP 4	Q7
13.	Entwurf einer Mustervereinbarung	AP 1	Q8
14.	Festlegung von Metadatenschema Version 2.0	AP 2	Q8
15.	Erhebung und Auswertung von Kennzahlen	AP 6	Q7-Q8
16.	Durchführung eines öffentlichen Projektabschluss-Workshops	AP 6	Q8

¹² Mögliche Use Cases mit Beispielen: <http://www.dlib.org/dlib/january12/lewis/01lewis.html> (Zugriff am 28.10.2014).

AP 1: Kommunikation mit den Allianz-Verlagen

Ziel: In Zusammenarbeit mit zwei oder drei Pilotverlagen sollen die Rahmenbedingungen für einen optimierten, systematischen Abruf der relevanten Publikationen von den Verlagsservern geschaffen werden. Ausgangspunkt sind dabei die bereits bestehenden Allianz- und Nationallizenzenverträge, die zwar Rechte festgelegt haben, nicht aber die Methode zur Wahrnehmung dieser Rechte.

Beschreibung und verwendete Methoden: Mit Sage Publications (<http://www.sagepub.com/home.nav>) und Karger (<http://www.karger.ch/>)¹³ konnten bereits zwei Verlage für die Mitarbeit in dem Projektvorhaben gewonnen werden. Die Zusagen beinhalten den intensiven Dialog über organisatorisch-technische Möglichkeiten der Datenlieferung seitens der Verlage. Gespräche werden auch geführt mit BMJ (<http://journals.bmj.com/>) (Stand 10/2014). Mit allen drei Verlagen hat die Bayerische Staatsbibliothek einen Allianz-Lizenzabschluss erzielt.

Mit diesen Pilotverlagen sollen Wege erarbeitet werden, wie das Publikationsaufkommen deutscher AutorInnen aus den zugriffsberechtigten öffentlich geförderten Wissenschaftseinrichtungen identifiziert, der jeweils relevanten Einrichtung zugeordnet und mit entsprechenden Metadaten ausgeliefert werden kann. Hierbei lassen sich ggf. auch etablierte Wege der Datenablieferung etwa an Europe PMC oder SCOAP3 nachnutzen. Es ist davon auszugehen, dass die Metadaten speziell mit Blick auf die Institution (Affiliation) der AutorInnen bei den Verlagen nicht optimiert bzw. durchgängig standardisiert sind. Deshalb sollten zunächst gewisse Unschärfen bezüglich der Vollständigkeit in Kauf genommen werden, da aufwändige Anpassungen der Daten auf Verlagsseite nicht realistisch sein dürften. Es muss vielmehr das primäre Ziel sein, Verlage überhaupt für die Kooperation zu gewinnen, zu der sie vertraglich nicht verpflichtet sind. AP 1 baut daher auf vertrauensbildende Maßnahmen, die in einer möglichen Hauptphase des Projektes auf weitere Verlage übertragbar wären. Auszuloten ist außerdem, welche Position die Verlage gegenüber einem für eine mögliche zweite Projektphase angedachten offenen Zentralrepositorium einnehmen.

Die erzielten Ergebnisse sollen dann Eingang finden in eine mit den Pilotverlagen abgestimmte Mustervereinbarung, die im Rahmen des Arbeitspakets entworfen wird. Diese soll in den Kontext der Verhandlung von Allianz-Lizenzen zurückgespielt werden, aber auch über diesen spezifischen Lizenzierungszusammenhang hinaus Verwendung finden.

Ergebnisse: 1. Klärung der Voraussetzungen für den automatisierten Abzug der relevanten Daten durch ein intermediäres Repositorium mit den Verlagen; 2. Vorlage einer Mustervereinbarung, die für künftige Verlagsverhandlungen verwendet werden kann.

Risiken: Ein Risiko besteht darin, dass mit den Verlagen keine Einigung über den erforderlichen Umfang der Zusammenarbeit erzielt werden kann. Folge: Die gewonnenen Verlage beenden die Zusammenarbeit. Auch könnten sich die durch die Verlage lieferbaren Daten als unzureichend erweisen (siehe auch AP 2 und AP 3 und dortige Maßnahmen zur Risikominimierung). Ein weiteres geringes Risiko besteht, dass bibliotheksseitig kein rechtssicherer und organisatorisch gangbarer Weg gefunden wird, flächendeckend das Einverständnis der WissenschaftlerInnen dahingehend zu erhalten, in ihrem Namen die Metadaten und Dokumente bei den Verlagen systematisch abgreifen zu können. Diese juristische Abklärung muss pro Institution gelöst werden. Im Rahmen des vorliegenden Projekts kann für diese Fragestellung nur sensibilisiert werden.

Maßnahmen zur Risikominimierung können Gespräche mit weiteren Verlagspartnern sein. Außerdem: die vertrauensbildende Kommunikation mit allen Parteien und die Entwicklung juristischer Empfehlungen für Vertragsanpassungen zur handhabbaren Einverständniseinholung bei der Autorenschaft vor Ort und zur langfristigen Etablierung des Dienstes. Bestenfalls wird perspektiv-

¹³ Der Schweizer Medizin-Verlag Karger wirkt bereits jetzt aktiv bei der Bereitstellung der Daten mit, die sich aus den Open-Access-Rechten der beteiligten Bibliotheken ergeben.

tivisch die Nutzung des Dienstes in die Grundsätze für den Erwerb DFG-geförderter überregionaler Lizenzen einbezogen.

AP 2: Entwicklung eines Datenschemas

Ziel: Entwicklung eines Metadatenschemas unter Berücksichtigung der Interessen und technischen Möglichkeiten aller Beteiligten.

Beschreibung und verwendete Methoden: Im AP 2 werden von den Pilotverlagen gelieferte Testdaten analysiert und in Abstimmung mit allen Projektteilnehmern (Verlage, Bibliotheken) und der DINI-AG Elektronisches Publizieren ein Metadatenchema entwickelt. Hierbei wird eine möglichst breite Nachnutzbarkeit (vgl. auch Abschnitt 2.4) und Konformität zu vorhandenen Standards, Schnittstellen und Ablieverfahren bei allen Beteiligten angestrebt. Sofern dies durch ein Metadatenchema nicht erreicht werden kann, werden zusätzliche Maßnahmen zur Datenkonvertierung, -normierung und -anreicherung diskutiert und in Arbeitspaket 3 abgebildet.

Dabei gilt es, verschiedene Perspektiven zu betrachten und pragmatisch miteinander zu verbinden:

- 1. Verlagsperspektive.** Laut Auskunft der DNB (Bereich Erwerbung Netzpublikationen) liefern die meisten Verlage ihre Publikationen im ONIX-Metadatenformat. Daneben unterstützt die Nationalbibliothek als Ablieferungsformate MARCXML und XMetaDissPlus.¹⁴ In der Mehrzahl unterstützen Verlage an Dublin Core (DC) angelegte Formate.

Verlag/ Produkt	Metadatenstandard	OAI-PMH	Informationen/Details gem. folgender URL
Oxford Journals	DC, MARC 21, NML, PRISM	Ja	http://www.oxfordjournals.org/en/help/tech-info/standards.html
IOPscience	DC, NML	Ja	http://iopscience.iop.org/page/STACKS
Karger Journals	MARC, ONIX	Nein	www.karger.com/WebMaterial/ShowFile/12157
Cambridge Journals	DC (z.T. modifiziert)	Ja ¹⁵	http://journals.cambridge.org/action/stream?pagelId=9048&level=2&sessionId=D95EA6D980A3297454B3B21080784CD1.journals
AIP Journals	NML	Nein	n/a (Auskunft von Rich O'Keeffe, AIP)
BMJ Journals	NML	Nein ¹⁶	http://highwire.stanford.edu/librarians/hwcurrent.dtl und http://highwire.stanford.edu/sru/sru_search.dtl
De Gruyter	MARC21, ONIX		http://www.degruyter.com/page/835 , Auskunft von Maren Brodersen, De Gruyter

Tabelle 1: Ermittelte Metadatenstandards von Verlagen (Stand 9/2014)

- 2. Bibliotheksperspektive.** Die Metadaten sollen im Idealfall auch in Bibliothekskataloge bzw. in Discovery-Systeme integriert werden können. Ein wichtiges Anliegen ist hier die hinreichende Tiefe der Metadaten, d. h. sie sollen die Komplexität und die Beziehungen eines Werkes abbilden (z.B. Beziehungen zwischen Artikel, AutorInnen der Artikel, Titel der Zeitschrift, dem jeweiligen Zeitschriftenheft etc.). In Hinblick auf die zukünftige Rolle der Bibliothek als Datenabnehmer muss auch die Zuordnung bedacht werden.¹⁷ Dafür muss zusätzlich die Affiliation (zugehörige Hochschule, wissenschaftliche Einrichtung, Institution o.ä.) der AutorInnen in den Metadaten verzeichnet werden, damit entsprechend selektiert werden kann.
- 3. Perspektive der Repositoriensoftware.** Gleichzeitig müssen die Systeme betrachtet werden. Laut dem 2014 'Census of Open Access Repositories' wird in Deutschland, Österreich und der Schweiz primär die Software OPUS verwendet (53,3 %), dann folgen EPrints (15,1 %) und

¹⁴ Vgl. <http://d-nb.info/1027296637/34>, S. 6 (Zugriff am 28.10.2014).

¹⁵ Lt. <http://journals.cambridge.org/action/stream?pagelId=2648&level=2&sessionId=D95EA6D980A3297454B3B21080784CD1.journals#3528> (Zugriff am 28.10.2014).

¹⁶ Der durch BMJ beauftragte Dienstleister Highwire nutzt stattdessen SRU (Search/Retrieve via URL).

¹⁷ Siehe hierzu auch die laufende Diskussion zu „JATS 4 Reuse“: <http://www.openaccessweek.org/profiles/blogs/jats-4-reuse-working-group-launch-first-public-event> (Zugriff am 28.10.2014).

DSpace (7,2 %).¹⁸ Weltweit ist DSpace die am weitesten verbreitete Repositoriensoftware (1.161) gegenüber EPrints (380) und OPUS (70).¹⁹ Als Schnittstelle kann OAI-PMH vorausgesetzt werden, auch SWORD wird häufig (EPrints, DSpace; nicht jedoch OPUS) unterstützt. Als Datenformat wird in der Regel Dublin Core (DC) verwandt (oder DC angelehnte Formate). Auch die bekannten Repositorien Europe PMC und PubMed Central (PMC) kodieren nach DC.²⁰ Im Rahmen von DRIVER und OpenAIRE wurde zudem ein kontrolliertes Vokabular für europäische Repositorien namens info:eu-repo entwickelt,²¹ dem ebenfalls Beachtung geschenkt werden muss. Die Antragsteller stehen dazu mit der zuständigen Interest Group „Controlled Vocabularies for Repository Assets“ der Confederation of Open Access Repositories (COAR) in Kontakt.²²

Zur Vernetzung des Projektes DeepGreen ist für das zweite Quartal ein zweitägiger Workshop geplant, in dessen Rahmen mit geladenen Experten aus Wissenschaft, Informationsinfrastruktur und Verlagen die Anforderungen an das Metadatenschema (AP 1) und den Workflow (AP 3, AP 4 und AP 5) diskutiert werden sollen.

Der Workshop dient auch als Plattform, um Szenarien für die Weiterentwicklung der DeepGreen-Infrastruktur zu diskutieren. Ziel ist es, die Lücken bestehender Protokolle (Workflows) und Standards (Metadaten) zu benennen und Szenarien zu konzipieren. Auch Themen wie AutorInnen- und Institutionenidentifikation sollen thematisiert werden. Die Durchführung des Workshops ist im Dialog mit der DINI-AG Elektronisches Publizieren geplant. Vertreter der DINI AG Forschungsinformationssysteme (FIS) und dem Kompetenzzentrum Interoperable Metadaten (KIM) sowie die relevanten Arbeitsgruppen der Schwerpunktinitiative "Digitale Information" sollen eingebunden werden. Die Ergebnisse des Workshops werden in einem öffentlichen Report dokumentiert.

Ergebnisse: Festlegung und Veröffentlichung des Metadatenkonzepts Version 1; Überarbeitung und Anpassung des Metadatenkonzepts und Veröffentlichung der Version 2.

Risiken: Für den Erfolg des Vorhabens ist wesentlich, dass die Interessen aller Stakeholder beachtet und möglichst gut ausbalanciert werden.

Maßnahmen zur Risikominimierung: Risikominimierend wirkt die Ausrichtung an internationalen Datenablieferungsverfahren. Es kann weder davon ausgegangen werden, dass Verlage neue Schnittstellen verwenden noch dass sie ein neues Metadatenschema entwickeln. Dies gilt auch dann, wenn wissenschaftlich andere Standards wünschenswert wären. Entsprechend müssen ggf. Mappings bzw. Konvertierungsmaßnahmen geplant und berücksichtigt werden, um den kleinsten gemeinsamen Nenner aller Parteien zu bedienen.

¹⁸ Repository-Software gem. 2014 Census, vgl. http://repositoryranking.org/?page_id=664 (Zugriff am 28.10.2014).

¹⁹ OpenDOAR: <http://www.opendoar.org/> (Zugriff am 28.10.2014).

²⁰ Zu Metadatenformaten und Schnittstellen für OPUS 4: http://www.kobv.de/fileadmin/opus/download/opus_dokumentation_de.pdf; für EPrints: http://wiki.eprints.org/w/Repository_configuration; für DSpace: <https://wiki.duraspace.org/display/DSDOC4x/Metadata+and+Bitstream+Format+Registries>, <https://wiki.duraspace.org/display/DSDOC4x/Functional+Overview>. Die Repositorien Europe PMC und PubMed Central (PMC) kodieren nach Dublin Core (DC), sie akzeptieren dazu auch NML (NewGRF Meta Language). (Zugriffe jeweils am 28.10.2014).

²¹ Vgl. <https://wiki.surfnet.nl/display/standards/info-eu-repo> (Zugriff am 28.10.2014). Zu OpenAIRE siehe auch: <https://www.openaire.eu/guides/rep-man-toolkit/guidelines-update-for-repository-managers> (Zugriff am 28.10.2014).

²² Vgl. <https://www.coar-repositories.org/activities/repository-interoperability/ig-controlled-vocabularies-for-repository-assets> (Zugriff am 28.10.2014).

AP 3: Anforderungsanalyse / Entwurf eines Gesamtworkflows

Ziel: Erheben und Sammeln der formalen und technischen Anforderungen; Konzeption eines Gesamtworkflows für DeepGreen nach Absprache mit Verlagen und Repositorienbetreibern.

Beschreibung und verwendete Methoden: In AP 3 werden anhand von konkreten Anwendungsfällen (Use Cases) der im Projekt beteiligten Repositorienbetreiber die praktischen Workflows für DeepGreen modelliert. Schließlich wird für die softwaretechnische Umsetzung ein Flussdiagramm entwickelt. Ausgangspunkt dafür sind erste von den Projektpartnern überlegte Skizzen zu technischen Bausteinen und Workflows (siehe Anhänge: C_DeepGreen_AnhangI_Ideenskizze_Komponenten; C_DeepGreen_AnhangII_Ideenskizze_Workflow). Ein zentraler Punkt ist das Rechtemanagement. Es muss ein Verfahren gefunden werden, mit dem für alle Parteien dokumentiert wird, welche Bibliothek welche Dokumente abgerufen hat. Diese Informationen erleichtern die Arbeit für die Bibliothek, tragen aber im Wesentlichen für Verlage und Förderer zur gewünschten Transparenz bei. Idealtypisch kann für die betreffende Bibliothek die Auswahl der relevanten (neuen) Artikel optional um die bereits geholten reduziert werden.

Die Konzepte zur Datenprüfung und -validierung sind ein wichtiger Bestandteil zur vertrauensvollen Zusammenarbeit mit den Verlagen. Wann welche Daten für wen freigegeben werden, wird durch die Modellierung transparent. Daneben werden alle anderen Workflows abgebildet, z.B. die Konzeption zur Übernahme vom Verlag und die Übergabe zu den Repositorien. Entsprechend einer prototypischen Entwicklung mit einem agilen Ansatz erfolgen Anforderungsbeschreibung und Softwareumsetzung parallel.

Aus der Perspektive der bibliothekarischen Repositorien muss der Importvorgang modelliert werden. Hier stellt sich die Frage, wie bereits vorhandene Publikationen (Preprints / reine Metadaten) gefunden und mit den entsprechenden Verlagspublikationen zusammengeführt werden können.

Ergebnisse: Das Ergebnis dieses APs ist ein abgestimmter Katalog der fachlichen Anforderungen und eine technische Spezifikation (mit Priorisierungen nach MoSCoW²³). Die technische Spezifikation setzt auf offene Schnittstellen und Protokolle.

Risiken: Eine Gefahr besteht darin, dass Anforderungen und Erwartungen an die Implementierung formuliert werden, die mit den vorhandenen Ressourcen im gegebenen zeitlichen Rahmen nicht umgesetzt werden können. Entscheidend für den Projekterfolg wird dann eine kluge Priorisierung der Anforderung sein.

Maßnahmen zur Risikominimierung: Eingesetzt wird eine professionelle Projektleitung, die auch im Rahmen von Entwicklungsprojekten erfahren in Softwareentwicklungsprozessen ist und die Koordination übernimmt.

²³ Zur Erläuterung der MoSCoW-Methode siehe den entsprechenden Wikipedia-Beitrag: <http://de.wikipedia.org/wiki/MoSCoW-Priorisierung> (Zugriff am 28.10.2014).

AP 4: Implementierung eines Prototyps der Datendrehscheibe

Ziel: Aufbau und Implementation der technischen DeepGreen-Architektur.

Beschreibung und verwendete Methoden: Das AP 4 steht in enger Beziehung zu AP 3 und AP 2. Die Entwicklung erfolgt nach agilen Grundsätzen parallel zur Erstellung der Anforderungen. Grundsätzlich soll im Rahmen des Vorhabens keine Softwareneuentwicklung stattfinden, stattdessen soll bestehende Open-Source-Software genutzt und angepasst werden. Vermutlich eignet sich als technische Grundlage für die Datenanlieferung und Datenablieferung eine Repositoriumssoftware am ehesten (DSpace, Eprints, OPUS 4, Fedora, Invenio, o.ä.). Die Auswahlentscheidung für die zugrundeliegende Software ist von der Art der benötigten offenen Schnittstellen abhängig und für einen späteren Betrieb von der grundsätzlichen Skalierbarkeit der Anwendung. Die Planung sieht zunächst die Verlagsseite – Import von Metadaten und Volltexten (unter Berücksichtigung von Datenvalidierungs- und Berechtigungsmechanismen) – und im zweiten Schritt die Abgabe von Daten und Dateien an die projektbeteiligten Repositorien vor. Im Dialog mit den Verlagspartnern müssen vermutlich (analog zu Europe PMC und SCOAP3) unterschiedliche Ablieferungswege zur Verfügung gestellt werden, dabei müssen auch notwendige Erweiterungen für die einzelnen abnehmenden Repositorien überlegt werden. Favorisiert wird ein agiler, testgetriebener Entwicklungsansatz, der sich der bestehenden KOBV-Entwicklungsinfrastruktur einfügt.

Ergebnisse: Die technische Architektur wird in drei Versionen (Prototyp Version 1 mit den grundlegenden technischen Bausteinen; Prototyp Version 2 nach Tests mit Erweiterungen und Schnittstellen; nach Tests mit den Repositorien erscheint Prototyp Version 3) entwickelt. Die Veröffentlichung benötigter Erweiterungen erfolgt auf einschlägigen Plattformen wie GitHub.

Risiken: Für die Akzeptanz und den Erfolg entscheidend ist die Auswahl von international üblichen Standards und Schnittstellen.

Maßnahmen zur Risikominimierung: Entscheidend hierfür wird nicht zuletzt die Gewinnung eines weiteren geeigneten Projektmitarbeiters sein. Dieser sollte idealerweise fachlich erfahren mit bibliothekarischen Geschäftsgängen, bibliographischen Datenstrukturen und Schnittstellen und gleichzeitig ein guter Softwareentwickler sein. Um Code-Reviews durchzuführen und Fehler schon in der prototypischen Entwicklung zu minimieren, ist der/die ProjektmitarbeiterIn in ein Entwicklerteam einzubetten, das aus Eigenleistung aus der Entwicklerpartnerschaft zwischen BVB und KOBV finanziert werden kann.

AP 5: Implementierung des Workflows in den beteiligten Repositorien

Ziel: Erfolgreiche Übernahme von Metadaten und Volltexten in beteiligte Repositorien.

Beschreibung und verwendete Methoden: Das in AP 3 erstellte Konzept zur Übergabe der Metadaten und Dokumente an die Institutionen und deren Übernahme in die beteiligten Repositorien sollen in AP5 realisiert und erprobt werden. Übergabe und Übernahme sind in drei Schritte aufzuteilen: Zunächst müssen die Daten identifiziert werden, die die jeweilige Institution prinzipiell übernehmen könnte, d.h. Daten, die von den Verlagen abgeliefert wurden und für deren Übernahme und Publikation die Institution über entsprechende Rechte verfügt; anschließend muss innerhalb des Repositoriums der Institution geprüft werden, welche Datensätze bereits vorhanden sind und welche übernommen werden müssen. Dabei können gegebenenfalls vorhandene Preprints um Verlagsversionen und Metadaten vorhandener Publikationen um Quellenangaben ergänzt werden; im letzten Schritt werden die Daten der noch nicht im Repository vorhandenen Publikationen übernommen, wobei auf die richtige Zuordnung der übernommenen Daten innerhalb des Zielrepositorys geachtet werden muss und eine manuelle Kontrolle der übernommenen Daten zu ermöglichen ist. Während der erste Schritt zur Datenidentifikation auf Seiten von DeepGreen realisiert

werden kann, erfordern die anderen Schritte entsprechende Erweiterungen der Repositoriensoftware (DSpace, OPUS 4, eSciDoc/PubMan).

Ergebnisse: Je eine Erweiterung für die Repositoriensoftware DSpace, OPUS 4, und eSciDoc/PubMan zur weitgehend automatisierten Anbindung von DeepGreen und entsprechende Dokumentationen. Erweiterungen und Dokumentation werden über gängige Plattformen, z.B. GitHub, bereitgestellt.

Risiken: Schwierig abschätzbar ist der Aufwand für die Implementierung der Importschnittstellen und für eine Dublettenprüfung in den beteiligten Repositorien. Eine Dublettenprüfung gibt es bislang in keiner Repositoriensoftware.

Maßnahmen zur Risikominimierung: Für die Importroutinen sollen Standardschnittstellen (wie OAI oder SWORD) zur Anwendung kommen, auch für die Dublettenprüfung soll bei den Metadaten auf Standards gesetzt werden (siehe AP 2).

AP 6: Erhebung von Kennzahlen (Erfolgsmessung)

Ziel: Es werden nachprüfbare Kennzahlen erhoben, anhand derer der Erfolg von DeepGreen nach Ablauf der Pilotförderphase ermittelt werden kann. Die Entscheidung über einen Eintritt in Phase 2 basiert auf dem nachweislichen Erfolg von Phase 1.

Beschreibung und verwendete Methoden:

1. Erfolgskriterium in Bezug auf verlagsseitige Projektpartner: Da der Fall eintreten kann, dass einer der gewonnenen Verlage im Laufe des Projekts die Zusammenarbeit beendet, werden parallel Verhandlungen mit weiteren Verlagen durchgeführt. Am Ende müssen mindestens **zwei Verlage** das Projekt begleitet haben und für eine Folgephase zur Verfügung stehen.
2. Erfolgskriterium in Hinblick auf die Zahl der verlagsseitig abgelieferten Allianz-Lizenz-Artikel: Von Seiten der teilnehmenden Verlage werden mindestens **85 %** ihrer gesamten, unter Allianz-Lizenzen fallenden **Zeitschriftenartikel** in DeepGreen eingespielt.
3. Erfolgskriterium in Hinblick auf die testweise Ablieferung an institutionelle Repositorien: Wiederum **85%** der unter 2. in DeepGreen eingespielten Artikel mit Autoren aus den Einrichtungen der FAU Erlangen-Nürnberg, TU Berlin und GFZ Potsdam **können in das jeweilig zugehörige institutionelle Repository erfolgreich eingespielt werden.**

Sollten aus zwingenden technischen Gründen die Anzahl der Artikel bei Erfolgskriterium 2 und 3 reduziert werden müssen und die Antragsteller dennoch vom Erfolg des Projekts überzeugt sein, wird die tatsächliche Zahl eingespielter Artikel und die technische Begründung der DFG vorgelegt, so dass diese auch im Hinblick auf den Nachantrag für die zweite Projektphase über den Erfolg entscheiden kann.

Schließlich beschäftigt sich AP 6 mit der Übertragbarkeit von DeepGreen auf andere Kontexte: Insbesondere steht die Prüfung einer Übertragbarkeit auf Konsortialverträge außerhalb der DFG-Förderung im Fokus. Vereinzelt werden bei den in der Arbeitsgemeinschaft Deutscher, Österreichischer und Schweizer Konsortien (GASCO) vertretenen Einkaufsgemeinschaften bereits jetzt Verhandlungen über Zweitveröffentlichungsrechte geführt. DeepGreen kann dafür Handlungsoptionen aufzeigen und perspektivisch eine technisch-organisatorische Lösung zur optimierten Wahrnehmung der Zweitveröffentlichungsrechte anbieten. Die Übertragbarkeit der entwickelten Workflows über den engeren Kontext der bereits verhandelten Allianz-Lizenzen hinaus kann ebenfalls als Benchmark-Kriterium gelten.

Zum Abschluss der ersten Projektphase wird im achten Quartal ein eintägiger öffentlicher Workshop organisiert. Über die Vorstellung der entwickelten Infrastruktur hinaus sollen dort Visionen für

die Weiterentwicklung des Vorhabens erarbeitet werden. Zielgruppe sind Vertreter aus Wissenschaft, Informationsinfrastruktur und Verlagen.

Ergebnisse: Bei Projekterfolg steht von Seiten der teilnehmenden Verlage das gesamte, unter Allianz-Lizenzen fallende Verlagserzeugnis unter DeepGreen zur Verfügung (minus max. 15%, da ein komplett verlustfreier Übertrag erfahrungsgemäß unrealistisch ist). Den am Projekt teilnehmenden Repositorienbetreibern steht eine Erweiterung für die jeweilige Repositoriensoftware zur Verfügung, mit der sie wiederum mindestens 85% der in Frage kommenden Artikel ihrer Einrichtung ins institutionelle Repository importieren können.

Risiken: Ein Risiko liegt darin, dass die Gesamtsumme der Allianz-Lizenz-Artikel nicht zuverlässig ermittelt werden kann (siehe Erfolgskriterium 2).

Maßnahmen zur Risikominimierung: Falls sich diese Zahl nicht ermitteln lässt, kann von Stichproben (Artikel durchschnittlich pro Heft) hochgerechnet werden.

2.4 Maßnahmen zur Erfüllung der Förderbedingungen / Umgang mit Projektergebnissen

Wie im Abschnitt 1 beschrieben, ermitteln die Antragsteller die einschlägigen Standards auf Seiten von Verlagen, Bibliotheken und Repositorien national und international und nehmen sie in ihre Planungen auf, um einen möglichst breit ansprechenden, nachnutzbaren und zukunftsfähigen Dienst zu entwickeln.

Da das Projekt an einem Bibliotheksverbund (KOBV) angesiedelt ist, sind die Voraussetzungen für die langfristige Pflege des Dienstes und des Informationsangebotes nach Auslaufen der Förderung potentiell gegeben. Die technische Infrastruktur von DeepGreen würde sich komplementär zu dem Angebot "Opus- und Archivierungsdienste"²⁴ an der KOBV-Zentrale anbieten.

Die technische Plattform des Dienstes wird Open Source entwickelt und steht allen Nachnutzern frei zur Verfügung. Die aus dem Projekt resultierenden Publikationen werden in einem oder mehreren Repositorien der Antragsteller (bspw. auf OPUS FAU oder im DSpace-Repository der TU Berlin) Open Access veröffentlicht, die Entwicklungen und dazugehörige Dokumentationen werden auf einschlägigen Plattformen wie GitHub bereitgestellt.

Die nachhaltige Vernetzung des Projektes innerhalb der Fachcommunity wird von allen Projektpartnern unter Koordination der Projektleitung KOBV vorangetrieben. Geplant sind Rückkopplungen der Projektergebnisse zu diversen fachlichen Zusammenschlüssen wie der ICOLC, der DINI-AG Elektronisches Publizieren und der Confederation of Open Access Repositories (COAR), sowie die Präsentation bei Anwendern (Software-Anwendertreffen, Workshops, o.a.) und dem breiten Fachpublikum (Open-Access-Tage, Bibliothekartag).

3 Literaturverzeichnis

Keines

²⁴ Vgl. http://www.kobv.de/infos_fuer_bibliotheken/services_fuer_bibliotheken/opus_archivierung/ (Zugriff am 28.10.2014).