




KATHRIN BEHRENS^I, ROLAND BERTELMANN^{I,II}, JULIA
BOLTZE^{2,III}, KLAUS CEYNOWA^{I,IV}, JÜRGEN CHRISTOF^{3,V},
THOMAS DIERKES^{4,III}, JULIA ALEXANDRA
GOLTZ-FELLGIEBEL^{5,III}, MATTHIAS GROSS^{IV}, MICHAELA
HAMMERL^I, MARSA HAOUA^V, PETRA HEERMANN^{VI},
CORNELIA HOFFMANN^{VI}, TOBIAS HÖHNOW^{VII},
THORSTEN KOCH^{6,III}, MONIKA KUBEREK^{7,V}, HEINZ
PAMPEL^{8,II}, MARKUS PUTNINGS^{9,VI}, BEATE RUSCH^{10III},
HILDEGARD SCHÄFFLER^I, JENS SCHWIDDER^{11III},
KONSTANZE SÖLLNER^{12,VI}, HEDDA STAUB^I, EIKE
WANNICK^{II}


DeepGreen – Open Access Transformation Eine Handreichung für institutionelle Repositorien


Version 1.0


¹  0000-0002-5588-0290

²  0000-0002-0819-4271


³  0000-0001-8642-0425


⁴  0000-0001-6222-0592


⁵  0000-0001-9834-2268


⁶  0000-0002-1967-0077


⁷  0000-0002-1672-5271

⁸  0000-0003-3334-2771

⁹  0000-0002-6014-9048

¹⁰  0000-0001-7664-4097

¹¹  0000-0002-7476-3997

¹²  0000-0002-6263-7846

^I Bayerische Staatsbibliothek (BSB)

^{II} Helmholtz-Gemeinschaft, Helmholtz Open Science Koordinationsbüro

^{III} Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV)

^{IV} Bibliotheksverbund Bayern (BVB)

^V Technische Universität Berlin, Universitätsbibliothek (TUB)

^{VI} Friedrich-Alexander-Universität Erlangen-Nürnberg, Universitätsbibliothek (FAU)

^{VII} Deutsches GeoForschungsZentrum GFZ

Zuse Institute Berlin
Takustr. 7
14195 Berlin
Germany

Telephone: +49 30-84185-0
Telefax: +49 30-84185-125

E-mail: bibliothek@zib.de
URL: <https://www.zib.de>

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

DeepGreen – Open Access Transformation

Eine Handreichung für institutionelle Repositorien


Version 1.0


Kathrin Behrens	Roland Bertelmann ¹	Julia Boltze ²
Klaus Ceynowa	Jürgen Christof ³	Thomas Dierkes ⁴
Julia Alexandra Goltz-Fellgiebel ⁵	Matthias Groß	
Michaela Hammerl	Marsa Haoua	Petra Heermann
Cornelia Hoffmann	Tobias Höhnow	Thorsten Koch ⁶
Monika Kuberek ⁷	Heinz Pampel ⁸	Markus Putnings ⁹
Beate Rusch ¹⁰	Hildegard Schäffler	Jens Schwidder ¹¹
Konstanze Söllner ¹²	Hedda Staub	Eike Wannick


21. August 2019


Das DFG-Projekt DeepGreen zielt darauf ab, wissenschaftliche Publikationen, die Allianz-Lizenzen unterliegen und nach Embargofristen frei zugänglich gemacht werden dürfen, automatisch in Open-Access-Repositorien einzuspielen. Dazu wurde die DeepGreen-Datendrehscheibe entwickelt. Sie basiert auf dem Publications Router, einer Software des Joint Information Systems Committee (Jisc) aus Großbritannien, und wurde prototypisch an die Anforderungen von DeepGreen angepasst. In die Datendrehscheibe speisen Verlage in regelmäßigen Intervallen Notifikationen (Publikationen und deren Metadaten) ein. Anschließend werden berechnete Open-Access-Repositorien identifiziert und beliefert.


Um Repositorienbetreibern eine erfolgreiche Veröffentlichung der zugeordneten Artikel zu ermöglichen, werden in dieser Handreichung Empfehlungen und ein Workflow bereitgestellt. Das Dokument basiert auf den bisherigen Erfahrungen und der Zusammenarbeit der Projektpartner in der ersten DeepGreen-Projektphase. Weitergehende Informationen zum DeepGreen-Projekt finden sie im Literaturverzeichnis (Goltz-Fellgiebel & Putnings (2019), Koch et al. (2016), Putnings & Rusch (2016)).


¹  0000-0002-5588-0290


²  0000-0002-0819-4271


³  0000-0001-8642-0425


⁴  0000-0001-6222-0592


⁵  0000-0001-9834-2268


⁶  0000-0002-1967-0077


⁷  0000-0002-1672-5271

⁸  0000-0003-3334-2771

⁹  0000-0002-6014-9048

¹⁰  0000-0001-7664-4097

¹¹  0000-0002-7476-3997

¹²  0000-0002-6263-7846

1 Zweitveröffentlichungsrechte - Verantwortlichkeiten

Die Information, welchen Allianz-Lizenzen einzelne Artikel tatsächlich unterliegen, wird von der Datendrehscheibe bei der Elektronischen Zeitschriftenbibliothek (EZB) eingeholt. Die Embargofristen sind vertragsabhängig und damit institutionenspezifisch. Die Datendrehscheibe hat keine Möglichkeit der Überprüfung und Vermittlung dieser Fristen an eine Institution. Für die Umsetzung der Open-Access-Rechte in Allianz-Lizenzen und den damit ggf. verbundenen Embargofristen sind die Repositorienbetreiber verantwortlich.

Da die Zuordnung von Publikationen zu Institutionen durch die Affiliationsangaben der Institution erfolgt, sind "False Positive"-Lieferungen nicht auszuschließen. Deshalb müssen Publikationen zunächst in einem geschützten Bereich aufgenommen und gründlich überprüft werden, bevor sie veröffentlicht werden können. Repositorienbetreiber müssen zum einen sicherstellen, dass das Repository das Recht hat, die Publikationen zu veröffentlichen; zum anderen müssen die Embargofristen eingehalten werden. Die Einhaltung der Embargofristen und die Nutzung der korrekten Version obliegt den Repositorienbetreibern¹.

2 Technische Anforderungen

Anders als Bibliotheken bzw. Repositorien nutzen Verlage in der Regel Metadatenformate wie ONIX, DTD NLM/NISO JATS oder CrossRef-XML (Putnings & Rusch, 2016). Daher muss eine Interoperabilität zur Sicherstellung der Datenqualität und Erleichterung der üblichen Geschäftsgänge der Institutionen gewährleistet sein. Die Datendrehscheibe schafft eine Kompatibilität der Schnittstellen von verschiedenen Verlagen und unterschiedlichsten Endsystemen. Verlagsartikel können mit den endsystemspezifischen Metadaten der Verlage und vollständigem Text automatisch in die institutionellen Repositorien eingespielt werden.

Für diese erfolgreiche Datenabnahme der Repositorien werden folgende Schnittstellen von DeepGreen bereitgestellt:

- **SWORD** (Simple Webservice Offering Repository Deposit): Volltexte und Metadaten werden direkt von der Datendrehscheibe (dem SWORD-Client) an Repositorien übermittelt.
- **OAI-PMH** (Open Archives Initiative Protocol for Metadata Harvesting): Die Datendrehscheibe stellt nur Volltexte bereit, die von Repositorien gesammelt und weiterverarbeitet werden können.
- **Web-API**: Metadaten und Volltexte können in zwei sukzessiven Schritten via REST (Representational State Transfer) von der Datendrehscheibe abgeholt werden.

¹Siehe hierzu die Ausführungen in § 3 des Haftungsausschlusses, der zwischen dem ZIB als DeepGreen-Dienstebetreiber und dem an DeepGreen teilnehmenden Repositorienbetreiber abgeschlossen wird.

2.1 Vorbereitung der SWORD-Schnittstelle

Über die SWORD-Schnittstelle werden Daten übertragen. Sowohl SWORD v1 als auch SWORD v2 werden von der Datendrehscheibe bedient. Entscheiden sich die Repositorienbetreiber für die SWORD-Schnittstelle, müssen gewisse Vorkehrungen getroffen werden:

1. System vorbereiten

(a) Aktivierung der Schnittstelle

Repositorienbetreiber müssen die SWORD-Schnittstelle freischalten. Dabei sollte eine Beschreibung des Service mittels des Servicedokuments bereitgestellt werden, die auf jeden Fall eine Authentifizierung verlangt und unter einer URL abrufbar sein sollte. Des Weiteren muss ein Account für DeepGreen angelegt werden, der die Nutzung des SWORD-Services gestattet.

(b) Collection (Sammlung)

Repositorienbetreiber müssen sicherstellen, dass es eine geschützte Sammlung mit einem Freischaltungsworkflow für die Aufnahme von SWORD-Importen in ihrem System gibt.

(c) Metadaten-Einspeisung

Je nach verwendeter Software stellt DeepGreen folgende Metadatenformate für die Datenabnahme zur Verfügung: NLM/NISO JATS, RIOXX, DC, METS/MODS, DTD RSC, OPUS-XML und ESciDoc. Repositorienbetreiber müssen die erfolgreiche Aufnahme sämtlicher bibliographischer Angaben sicherstellen, wenn ihr System nicht eines der genannten Formate unterstützt, indem sie die gelieferten Metadaten korrekt in ihr systemeigenes Metadatenformat konvertieren. Gegebenenfalls müssen sie ein Mappingverfahren der importierten Metadaten auf interne Felder definieren. Hierfür empfiehlt sich die Nutzung von XSL-Transformationen. Solch ein Mapping-Skript kann hier gefunden werden: `sword-mods-ingest.xsl`.

2. Zugangsdaten einrichten

Nach erfolgreicher Anmeldung beim DeepGreen-Router können Institutionen ihre SWORD-Zugangsdaten auf `oa-deepgreen.de` hinterlegen. Folgende Angaben werden benötigt:

- (a) URL der geschützten Sammlung, in die DeepGreen-Publikationen hochgeladen werden
- (b) Benutzername des SWORD-Users
- (c) Passwort des SWORD-Users
- (d) URI des gewünschten Metadatenformats (Packaging preferences)

2.2 OAI-PMH

Mit OAI-PMH können Open-Access-Repositorien über einfache HTTP-Anfragen Metadaten aus den Beständen der Datendrehscheibe eingesammelt werden. Anschließend können diese Daten weiterverarbeitet und frei verfügbar gemacht werden. Das standardmäßig bereitgestellte Metadatenformat ist das Dublin-Core-Datenmodell (DC). Der Bezug der zu den Metadaten gehörenden Volltexte

selbst kann über die Web-API oder via SWORD bei den einzelnen Repositorien erfolgen.

Es ist zu beachten, dass Repositorienbetreiber lediglich Verlagspublikationen, die innerhalb der letzten 90 Tage von den Verlagen in die Datendrehscheibe eingespeist wurden, abrufen können.

Die Schnittstelle ist über <https://www.oadeepgreen.de/oaipmh/all?verb=identify> erreichbar. Weitere Informationen zur Implementierung sind in der OAI-PMH-Dokumentation von DeepGreen zu finden.

Hinweis: Auf <https://www.openarchives.org/pmh/tools/> ist eine Liste von nützlichen OAI-PMH Tools zu finden.

2.3 Web-API

Die Datendrehscheibe stellt eine REST-API bereit, mit der Publikationen abgeholt werden können. Durch die bereitgestellten Endpoints können Repositorienbetreiber in einem ersten Schritt Metadaten holen und in einem zweiten Schritt die dazugehörigen Volltexte abrufen.

Auch hier ist zu beachten, dass Repositorienbetreiber lediglich Verlagspublikationen, die innerhalb der letzten 90 Tage von den Verlagen in die Datendrehscheibe eingespeist wurden, abrufen können.

Die Basis-URL für GET- und POST-Anfragen an diese Schnittstelle ist <https://www.oa-deepgreen.de>. Weitere Informationen zu den Endpoints sind in der Web-API-Dokumentation von DeepGreen zu finden.

Zahlreiche kostenfreie REST-API Clients sowie Tools sind online verfügbar und können für die Einsammlung der Metadaten und Volltexte eingesetzt werden.

3 Empfohlener Workflow: Bearbeitungsschritte für die Archivierung der Notifikationen

ACHTUNG: Je nach gewählter Schnittstelle (SWORD, OAI oder API) unterscheidet sich der Workflow geringfügig. Bei OAI und API werden zunächst nur die Metadaten einer Publikation heruntergeladen. In einem zweiten Schritt können dann die Volltexte bezogen werden. Dadurch können Institutionen vorab entscheiden, ob die Publikation für sie relevant ist oder nicht.

Der folgende Absatz stellt Repositorienbetreibern praktische Workflowschritte vor (vgl. Abbildung 1: DeepGreen-Workflow auf Repositorienseite). Diese können zur effektiven Bearbeitung und zur Zweitveröffentlichung von wissenschaftlichen Verlagsartikeln nach dem Open-Access-Prinzip der Allianzlizenzen in Open-Access-Repositorien über die Datendrehscheibe genutzt werden.

1. Überprüfung der Affiliationen der Publikationen

Eine korrekte Lieferung der einzelnen Publikationen von Autor*innen der jeweiligen Einrichtungen ist nicht hundertprozentig sichergestellt. Aufgrund der individuellen Affiliation-Konfigurationsdatei der jeweiligen Institutionen lassen sich "False Positive"-Lieferungen nicht vermeiden. Deshalb muss die Zugehörigkeit der beteiligten Autor*innen zu ihren jewei-

ligen Institutionen in vollem Umfang verifiziert werden. Hierfür kann der DOI (engl. Digital Object Identifier) eines Artikels auf der offiziellen Webseite des zuständigen Verlags verwendet werden, um die Affiliationen der Autor*innen zu ermitteln.

2. **Erkennung von Dubletten**

Die Datendrehscheibe ist ein reines Push-Forward-System. Verlagspublikationen werden einfach an die berechtigten Repositorien weitergereicht. Um die redundante Aufnahme von Verlagspublikationen im System zu vermeiden, empfiehlt sich vorzugsweise ein automatisierter Dublettencheck. Repositorienbetreiber können dann vorab entscheiden, ob die Publikation im Repository akzeptiert oder abgelehnt wird. Die Erkennung von Dubletten sollte so früh wie möglich erfolgen, um unnötige Arbeitsgänge zu vermeiden. Für DSpace-Systeme zum Beispiel gibt es bereits einen solchen automatischen Dublettencheck.

3. **Korrektheit der Metadaten**

Es muss sichergestellt werden, dass die Metadaten auf ihre Korrektheit und Vollständigkeit geprüft werden. Besonderes Augenmerk muss auf folgende Metadatenangaben gelegt werden:

(a) **Embargo**

Embargofristen werden nicht in den Metadaten übermittelt, sondern lediglich als Hinweis in den Notifikationen mitgeliefert. Liegt für eine Publikation ein Embargo vor, muss dieses eingehalten werden. Eine Publikation darf erst nach Ablauf des Embargos veröffentlicht werden. Embargofristen unterscheiden sich je nach Vertrag und Verlagspublikation. Die EZB-Schnittstelle bietet eine erste Anlaufstelle für die Ermittlung der Fristen. Nichtsdestotrotz sollte die Prüfung des jeweiligen Vertragstextes erfolgen.

(b) **Lizenz**

Falls Nutzungsrechte einer Publikation im Zuge einer Allianz-Lizenz eingeräumt wurden, müssen diese eingehalten und in den Metadaten kenntlich gemacht werden. Jegliche unberechtigten Erweiterungen sollten vermieden werden. Die Autoren der Handreichung "Open-Access-Rechte in Allianz und Nationallizenzen" (Stöber (2012)) empfehlen folgende deutsche bzw. englische Formulierung:

“Dieser Beitrag ist mit Zustimmung des Rechteinhabers aufgrund einer (DFG-geförderten) Allianz- bzw. Nationallizenz frei zugänglich.”

“This publication is with permission of the rights owner freely accessible due to an Alliance licence and a national licence (funded by the DFG, German Research Foundation) respectively.”

(c) **Dewey-Dezimalklassifikation (engl. Dewey Decimal Classification, kurz DDC)**

Die DDC-Sachgruppen einer Publikation werden von Verlagen nicht in den Metadaten bereitgestellt. Diese Werte müssen auf verschiedenen Wegen ermittelt werden. Repositorienbetreiber haben zum Beispiel die Möglichkeit, diese Werte mittels einer Software wie GATE,

BASE usw. automatisch zu ermitteln und ihre Metadaten mit DDC-Notationen anzureichern.

(d) **Version der Publikation**

Welche archivierbare Version zur Zweitveröffentlichung verwendet wird, muss entsprechend gekennzeichnet werden.

4. **Zweitveröffentlichungsort (Collection)**

Da Publikationen zu Beginn erst in einem geschützten Bereich aufgenommen werden, sollten sie nach der Bearbeitung archiviert und gegebenenfalls in einen adäquaten Sammlungsbereich bzw. eine Collection geschoben werden.

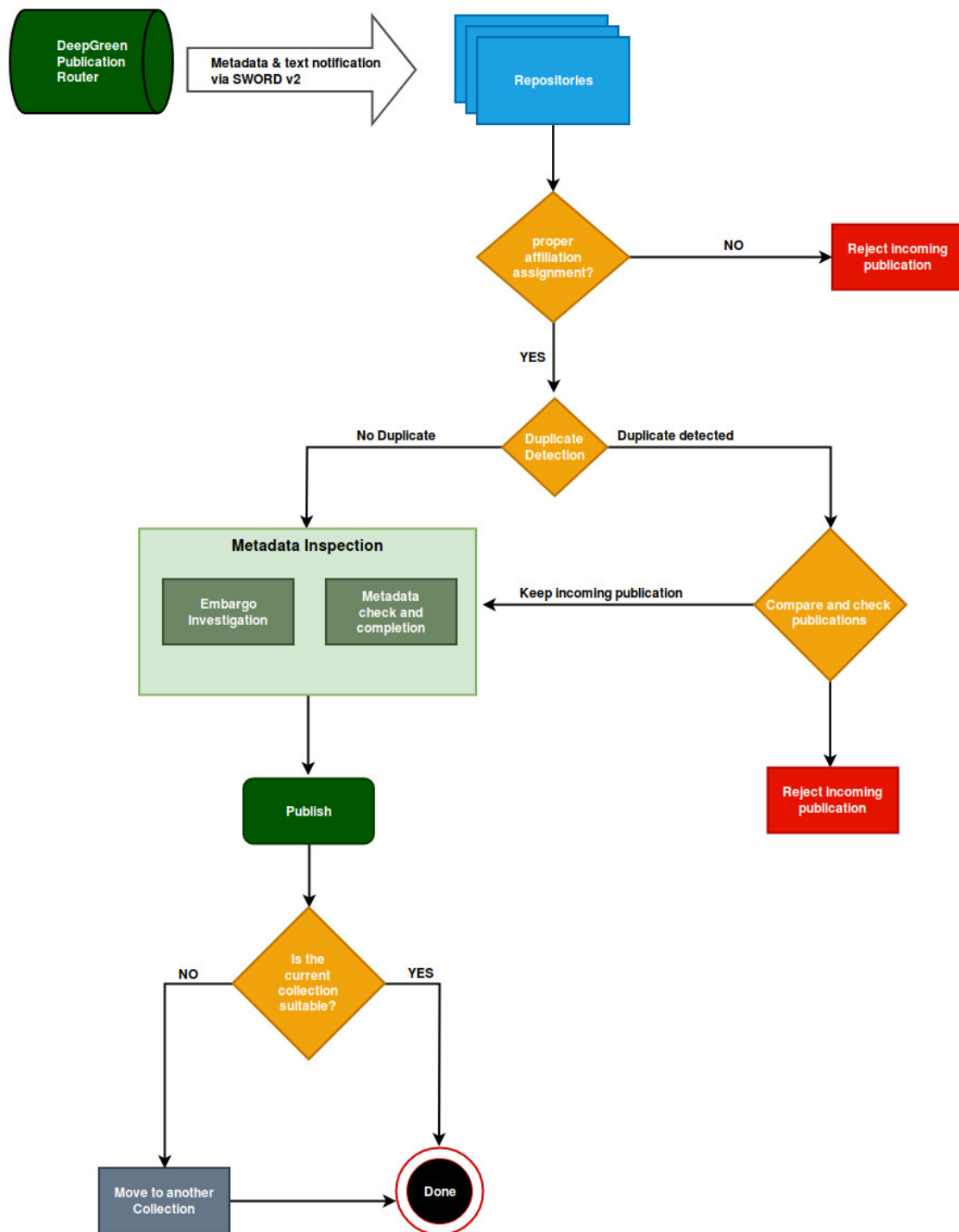


Abbildung 1: DeepGreen-Workflow auf Repositorienseite

Literatur

- [1] GOLTZ-FELLGIEBEL, J. A. ; PUTNINGS, M. : Open-Access-Transformation mit DeepGreen: Gemeinsam den (grünen) Schatz heben. In: *o-bib. Das offene Bibliotheksjournal* 6 (2019), Nr. 1, S. 1–11. – doi: <http://doi.org/10.5282/o-bib/2019H1S1-11>
- [2] KOCH, T. ; GRIEBEL, R. ; SÖLLNER, K. ; CHRISTOF, J. ; BERTELMANN, R. : DeepGreen - Entwicklung eines rechtssicheren Workflows zur effizienten Umsetzung der Open-Access-Komponente in den Allianz-Lizenzen für die Wissenschaft. ZIB-Report 15-58 (2016), S. 22. – urn:nbn:de:0297-zib-56799
- [3] PUTNINGS, M. ; RUSCH, B. : DeepGreen - Entwicklung eines rechtssicheren Workflows zur effizienten Umsetzung der Open-Access-Komponente in den Allianz-Lizenzen für die Wissenschaft. In: *o-bib. Das offene Bibliotheksjournal* 3 (2016), Nr. 4, S. 110–119. – doi: <http://doi.org/10.5282/o-bib/2016H4S110-119>
- [4] STÖBER, A. : Open-Access-Rechte in Allianz- und Nationallizenzen: Eine Handreichung für Repository-Manager, Bibliothekare und Autoren. Arbeitsgruppen „Nationale Lizenzierung“ und „Open Access“ der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen (Hrsg.). (2012), S. 17. – doi: <http://doi.org/10.2312/allianzoa.004>