

## Kongressbeiträge

# Open-Access-Transformation mit DeepGreen: Gemeinsam den (grünen) Schatz heben

Markus Putnings, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Universitätsbibliothek Erlangen-Nürnberg

Julia Alexandra Goltz-Fellgiebel, Zuse-Institut Berlin (ZIB), Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV)

### Zusammenfassung:

Das von der DFG seit 2016 geförderte Projekt DeepGreen will die Open-Access-Transformation der deutschen Wissenschaft unterstützen, indem Artikel, sofern lizenzrechtlich erlaubt, automatisiert in den grünen Weg von Open Access überführt werden. Dazu haben die Projektpartner – die Bibliotheksverbände Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV) und Bibliotheksverbund Bayern (BVB), die Bayerische Staatsbibliothek (BSB), die Universitätsbibliothek der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) und die Universitätsbibliothek der Technischen Universität Berlin (TU Berlin) sowie das Helmholtz Open Science Koordinationsbüro am Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum (GFZ) – prototypisch eine Datendrehscheibe auf Grundlage existierender Softwarebausteine entwickelt. Die beteiligten Verlage können hiermit Metadaten und Volltexte zyklisch über definierte Schnittstellen abliefern, die Daten werden anschließend rechtskonform an die dazu berechtigten institutionellen Repositorien weitergeleitet. Als Pilotpartner konnten S. Karger und SAGE Publications gewonnen werden, weitere Verlage beteiligten sich mit der Zusendung von Testdaten. DeepGreen hat sich in der Projektphase 2016-2017 auf die sogenannten Allianz-Lizenzen fokussiert, die seit 2011 zwischen Verlagen und Bibliotheken verhandelt wurden. Sie beinhalten spezifische Regelungen zum grünen Open Access, die einen enormen Mehrwert gegenüber den üblichen „Self-archiving policies“, also den Regelungen zur Zweitveröffentlichung bei Verlagen darstellen: Autorisierte Autorinnen und Autoren und deren Einrichtungen dürfen ihre Publikationen in der Regel in der publizierten PDF-Version nach verkürzten Embargofristen bzw. teils unmittelbar in ein Repository ihrer Wahl einstellen und öffentlich zugänglich machen. Praktisch wurde von dieser Möglichkeit bisher nur sehr eingeschränkt Gebrauch gemacht, DeepGreen soll dies perspektivisch ändern. Darüber hinaus erprobt DeepGreen in der zweiten Projektphase 2018-2020 die Ausweitung des Systems auf andere Lizenzmodelle und neue Datenabnehmer.

### Summary:

The DeepGreen project, funded by the German Science Foundation since 2016, aims to support the open access transformation of German science by automatically transferring articles into the green path of Open Access, if allowed by licence. Therefore the project partners – the library consortia Cooperative Library Network Berlin-Brandenburg (KOBV) and Bavarian Library Network (BVB), the Bavarian State Library (BSB), the university library of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), the university library of the Technische Universität Berlin (TU Berlin) and the Helmholtz Open Science Coordination Office at the German Research Center for Geosciences (GFZ) – have prototypically developed a data hub based on existing software components: Participating

publishers can provide metadata and deliver full texts cyclically via defined interfaces; data and full texts will be forwarded to authorized institutional repositories in accordance with legal requirements. As pilot partners, the publishers S. Karger and SAGE Publications were acquired, other publishers participated with delivering test data. In the project phase 2016-2017, DeepGreen focused on the so-called Alliance licenses, which have been negotiated between publishers and libraries since 2011. They contain specific regulations on green Open Access, which represent a huge added value compared to the usual "self-archiving policies" of publishers: Authorized authors and their institutions may generally put their publications (the published PDF-version) into a repository of their choice and make it publicly available directly or after a shortened embargo period. In practice, this option has not been used much so far, but DeepGreen aims to change this in the future. In addition, DeepGreen examines the possible extension of the system to other licensing models and new data reusers during the second project phase 2018-2020.

**Zitierfähiger Link (DOI):** <https://doi.org/10.5282/o-bib/2019H1S1-11>

**Autorenidentifikation:** Putnings, Markus: GND 1043244379

ORCID: <https://orcid.org/0000-0002-6014-9048>

Goltz-Fellgiebel, Julia Alexandra:

ORCID: <https://orcid.org/0000-0001-9834-2268>

**Schlagwörter:** Allianz-Lizenzen, Hybrides Open Access, Open-Access-Publikationen

## 1. Projektsteckbrief

Der 107. Deutsche Bibliothekartag stellte 2018 in Berlin die Frage, ob Open Access zum jetzigen Stand zukunftsfähig<sup>1</sup> sei: Die Referentinnen und Referenten bejahten dies, auch wenn Rafael Ball (Bibliothek der Eidgenössischen Technischen Hochschule Zürich, ETH) und Ulrich Herb (Saarländische Universitäts- und Landesbibliothek, SULB) manche Aspekte des Open-Access-Publizierens durchaus kritisch und kontrovers ansprachen.<sup>2</sup> Mit Blick auf den sogenannten grünen Weg des Open Access konstatierte etwa Herb: „Green Open Access ist zu zeitraubend, unattraktiv und umständlich.“<sup>3</sup> Julia Goltz-Fellgiebel (Kooperativer Bibliotheksverbund Berlin-Brandenburg, KOBV) und Markus Putnings (Universitätsbibliothek Erlangen-Nürnberg) stellten daraufhin das Projekt DeepGreen als einen möglichen Ansatzpunkt vor, um die benannten Probleme durch die im Projektkontext erarbeiteten Automatismen zu beheben oder zu reduzieren. Der nachfolgende Text stellt ihren Kongressbeitrag im Detail vor.

DeepGreen will wissenschaftliche Publikationen, die lizenzrechtskonform nach Ablauf eines Embargos zweitveröffentlicht werden dürfen, mit direkter Unterstützung der Verlage Open Access verfügbar machen, ohne den Autorinnen und Autoren zusätzlichen Aufwand zu verursachen. Dazu baut das

---

1 So auch der Titel der Session „Open Access – Zukunftsfähig?“ am 14.06.2018 von 09:00 - 10:30 Uhr im Themenkreis 5: „erwerben & lizenzieren“ des Deutschen Bibliothekartags 2018.

2 Vgl. Keuenhof, Maximilian: Open Access in der Kritik, *bibliotheksnews*, 14.06.2018, <<https://bibliotheksnews.com/2018/06/14/open-access-in-der-kritik/>>, Stand: 31.01.2019.

3 Herb, Ulrich: Open Access – Erfolge und nicht-intendierte Folgen, BIB OPUS Publikationsserver, 04.06.2018, <<https://nbn-resolving.org/urn:nbn:de:0290-opus4-36015>>.

Projekt eine vertragskonforme, automatisierte Datendrehscheibe für Metadaten und Volltexte auf, die die Verlagspublikationen an berechnigte Repositorien senden soll. Der manuelle Aufwand für die Selbstarchivierung wird damit für die Forschenden und die Institutionen minimiert. Auf der Projekt-homepage <https://deepgreen.kobv.de/> sind die verschiedenen Projektziele und -inhalte für Verlage und Repositorienbetreiber auf Deutsch und Englisch detailliert beschrieben, auch multimedial mit Videotutorials in Form von Screencasts.

Das Projekt wird durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen des Programms „Wissenschaftliche Literaturversorgung- und Informationssysteme (LIS)“, konkret über die Ausschreibung „Open-Access-Transformation“, gefördert. Die erste Förderphase lief vom 01.01.2016 bis zum 31.12.2017, die zweite Förderphase – wiederum für zwei Jahre bewilligt – startete am 01.08.2018 und läuft bis zum 31.07.2020.

Am Projekt beteiligt sind der KOBV, die Bayerische Staatsbibliothek (BSB) und der Bibliotheksverbund Bayern (BVB), die Universitätsbibliotheken der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) und der Technischen Universität Berlin (TU Berlin) sowie das Helmholtz Open Science Koordinationsbüro am Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum (GFZ).

Verlagsseitig erhält DeepGreen seit Projektbeginn 2016 Unterstützung von den Verlagen S. Karger und SAGE Publications, seit Sommer 2017 beteiligen sich auch die Verlage BMJ, De Gruyter und MDPI.

## 2. Hintergrund

Bibliotheken pflegen institutionelle Repositorien unter anderem mit dem Ziel, möglichst viele bzw. alle Veröffentlichungen der Mitarbeitenden der eigenen Institution nachzuweisen und/oder diese frei zugänglich zu machen. Problematisch ist hierbei in der Praxis weniger das Pflegen als das „Bepflegen“ der Repositorien, denn die Forschenden möchten möglichst wenig bis keine zusätzliche Arbeit damit haben. Zudem ist „[d]ie rechtliche Unsicherheit, insbesondere wenn es um Zweitveröffentlichungen auf dem Repository geht, [...] oft groß“.<sup>4</sup>

Das Einpflegen könnte in Teilen die Bibliothek als institutionelle Vertretung der Mitarbeitenden einer Einrichtung übernehmen, z.B. über eine entsprechende Servicestelle in der Bibliothek für die Forschenden vor Ort.<sup>5</sup> Die Rechtsproblematik bleibt allerdings bestehen, die rechtlichen Unsicherheiten und Aufwände verlagern sich nur: Die Bibliothek – insbesondere wenn keine Hochschulbibliografie vorliegt – kennt nicht alle Veröffentlichungen der Mitarbeitenden ihrer Institution, sie muss die Publikationsdaten pro Angehörigen der Einrichtung deshalb einzeln recherchieren und herunterladen,

---

4 Oberländer, Anja: Förderung von Open Access über institutionelle Infrastrukturen, insbesondere Repositorien, in: Söllner, Konstanze; Mittermaier, Bernhard (Hg.): Praxishandbuch Open Access (E-Book), Berlin; Boston 2017, S. 137-145, <<https://doi.org/10.1515/9783110494068-016>>, hier S. 144.

5 Vgl. Tobias, Regine: Die Quote kommt – Einwerbung von Open-Access-Publikationen durch nutzer-nahe Workflows im Repository, BIB OPUS Publikationsserver, 04.06.2018, <<https://nbn-resolving.org/urn:nbn:de:0290-opus4-35961>>.

anschließend rechtliche Abwägungen zur Zweitveröffentlichung treffen und die Daten wiederum ins Repositorium hochladen.

Der Gedanke, die Unterstützung der Verlage bei dieser Aufgabe anzufordern, liegt entsprechend nah, insbesondere da die Volltexte, die bibliografischen Metadaten und in großen Teilen auch die Angaben zur Affiliation der Autorinnen und Autoren sowie zur Rechtelage auf Verlagsseite bereits erfasst und vorgehalten werden. Viele Verlage unterstützen zudem die Zweitveröffentlichung als langzeitarchivierte Version der wissenschaftlichen Arbeit – teils eher nominell über Policies, Verträge und Lizenzen, teils auch konkret mittels automatischer Bereitstellung über SWORD<sup>6</sup>, wie es z.B. der Open-Access-Verlag MDPI anbietet.<sup>7</sup>

Auch im Hinblick auf die Umsetzung diverser Open-Access-Strategien in Bund, Ländern und Forschungseinrichtungen<sup>8</sup> scheint die Zusammenarbeit zwischen Verlagen und Informationseinrichtungen eine interessante Idee. Bestehende Maßnahmen bieten hier Anknüpfungsfelder: So wurden und werden auf bundesweiter Ebene DFG-geförderte Allianz-Lizenzen und Nationallizenzen mit weitgehend grünen Open-Access-Rechten verhandelt. In der Praxis werden diese aber bislang kaum genutzt,<sup>9</sup> da sie zum einen den Forschenden meist unbekannt sind, zum anderen die Wahrnehmung dieses Rechts im Sinne eines Uploads in Repositorien – wie auch von Herb erwähnt – für Forschende und die sie vertretenden Institutionen mit Zeit- und Arbeitsaufwand verbunden ist.

### 3. Kooperation mit den Verlagen

Dass die Verlage grundlegend willens sind, die benötigten Metadaten und Volltexte (PDF) zu liefern, sieht man an bestehenden Kooperationen, z.B. anhand des von Jisc angebotenen Publications Router, mit dem Veröffentlichungen u.a. der Verlage Frontiers, Hindawi, MDPI und BMJ an teilnehmende Repositorien in Großbritannien übertragen werden,<sup>10</sup> oder anhand der regelmäßigen Datenlieferungen an PubMed Central wie auch an nationale Pflichtabgabestellen. Die grundsätzliche Bereitschaft der Verlage, eng mit Bibliotheken zu kooperieren, spiegelt sich ebenfalls im Projekt DeepGreen wider: Schon in der Antragsphase fanden sich mit S. Karger und SAGE Publications kooperative Projektpartner, darüber hinaus lieferten die European Mathematical Society, De Gruyter, Oxford University Press, die Royal Society of Chemistry und MDPI im Laufe des Projekts Testdaten. Diese Unterstützung reflektiert auch die Grundsätze für den Erwerb DFG-geförderter überregionaler Lizenzen, denen sich die genannten Allianz-Lizenzverlage verschrieben haben:

6 SWORD (Simple Web-service Offering Repository Deposit) beschreibt ein Webprotokoll zum vereinfachten, automatisierbaren Einspielen von Dokumenten in Repositorien zum Zwecke der Veröffentlichung und Langzeitarchivierung, vgl. About SWORD. What is SWORD?, <sword />, 01.05.2018, <<http://swordapp.org/about/>>, Stand: 31.01.2019.

7 Vgl. Institutional Open Access Program (IOAP). Institutional Repository Deposits, MDPI, 01.01.2019, <<http://www.mdpi.com/about/ioap>>, Stand: 31.01.2019.

8 Vgl. Der freie Zugang zu wissenschaftlicher Information. Positionen, open-access.net, 31.01.2019, <<https://open-access.net/informationen-zu-open-access/positionen/>>, Stand: 31.01.2019.

9 Vgl. Schäffler, Hildegard: Open Access in konsortialer Perspektive, in: Söllner, Konstanze; Mittermaier, Bernhard (Hg.): Praxishandbuch Open Access (E-Book), Berlin; Boston 2017, S. 197-204, <<https://doi.org/10.1515/9783110494068-023>>.

10 Vgl. Publications Router. Current content providers, Jisc, 01.04.2018, <<https://pubrouter.jisc.ac.uk/about/providerlist/>>, Stand: 31.01.2019.

„Der Anbieter erklärt sich bereit, autorisierte Einrichtungen bei der Identifizierung und Lieferung relevanter Artikeldaten und Volltexte, einschließlich der digitalen Objekte, ohne Mehrkosten beratend und technisch zu unterstützen. Das kann z.B. darin bestehen, die Volltexte inkl. der Metadaten in einem gängigen Format bzw. nach solchen Standards bereit zu stellen, die ein Einspielen in Repositorien erleichtern. Begrüßenswert wäre zudem, dass der Anbieter es selbst übernimmt, Artikel von Autoren aus autorisierten Einrichtungen in ein vereinbartes Repository einzupflegen, z.B. über eine SWORD-Schnittstelle.“<sup>11</sup>

Um entsprechende Datenlieferungen seitens der Verlage erfolgreich an berechnigte Repositorien zu ermöglichen, muss die technische Implementierung der Datenannahme und -weitergabe

- die (ablauf-)technischen Voraussetzungen der Verlage berücksichtigen,
- die Inhalte der Lizenzverträge rechtssicher abbilden, und
- auf die lokalen Voraussetzungen der Daten empfangenen Repositorien Rücksicht nehmen (z.B. über standardisierte und im Bedarfsfall auch vorab zu konvertierende Importformate, über Schnittstellen und Protokolle).

DeepGreen hat hierzu in der ersten Projektphase intensiv mit den beteiligten Verlagen und Repositorien sowie unterstützenden Einrichtungen wie der Elektronischen Zeitschriftenbibliothek (EZB) diskutiert und folgende Fragen geklärt:

- Welche Daten (z.B. Preprints, Postprints, Verlagsversion; fokussierte Vertragsinhalte) können/wollen Verlage liefern?
- Welche Qualität haben die Affiliationsangaben in den Verlagsdaten?
- Welche Schnittstellen können Verlage bedienen?
- Welche Bibliothek nimmt an welchem (Allianz-/Nationallizenz-)Vertrag teil?
- Welche Zeitschriftentitel umfasst der (Allianz-/Nationallizenz-)Vertrag?
- Welche Zeitschriftenjahrgänge umfasst der (Allianz-/Nationallizenz-)Vertrag?
- Welche Formate, Schnittstellen und Standards können Verlage/Bibliotheken/Repositorien bedienen?<sup>12</sup>

## 4. Die technische Umsetzung

Als technische Lösung wurde ein weitgehend automatisierter Workflow über eine zentrale Datendreh-scheibe unter Einbeziehung der Verlage und bereits bestehender Softwarekomponenten konzipiert. Konkret wurde die Basisversion des bereits erwähnten Publications Router von Jisc herangezogen, der ein ähnliches Konzept einer Datendreh-scheibe und -auslieferung verfolgt und den bereits mehrere Verlage nutzen. Die ergänzten bzw. abgeänderten Funktionalitäten zur Jisc Software sind offen

---

11 12.181 – Grundsätze für den Erwerb DFG-geförderter überregionaler Lizenzen (Allianz-Lizenzen), DFG, 01.03.2015, <[www.dfg.de/formulare/12\\_181/](http://www.dfg.de/formulare/12_181/)>, Stand: 31.01.2019.

12 Vgl. Becker, Pascal-Nicolas; Bertelmann, Roland; Ceynowa, Klaus u. a.: Questionnaire for effective exchange of bibliographic metadata – current status of publishing houses, Dokumentenserver des Zuse Institute Berlin (ZIB), 29.09.2016, <<http://nbn-resolving.de/urn:nbn:de:0297-zib-60419>>.

in GitHub dokumentiert,<sup>13</sup> so mussten bei DeepGreen zusätzlich die in Deutschland gebräuchlichen Repositoriensoftwaretypen OPUS 4<sup>14</sup> und DSpace<sup>15</sup> berücksichtigt, die Konvertierung des Metadatenformats in die jeweiligen Importformate der Repositorien (z.B. OPUS-XML-Importformat<sup>16</sup>) realisiert und eine Schnittstelle zum Abgleich der Lizenzrechte über die EZB integriert werden.

Bei der Datenlieferung, -zuordnung und -zustellung mit DeepGreen steht ein benutzerzentrierter Service im Mittelpunkt; so sollten sowohl Verlage als auch Repositorien entsprechende Konten mit Profilinformationen, z.B. zur SWORD-Schnittstelle, zum lizenzierten Verlagsausschnitt und mit detaillierten Affiliationsangaben, einrichten können. Die Daten werden seitens DeepGreen gemäß dem Verlagskontenprofil angenommen, gegen Kontenprofile von Bibliotheken abgeglichen und gemäß lizenzrechtlicher Zuordnung bereitgestellt.

Bei DeepGreen wird hierbei eine zweistufige Berechtigungslogik eingesetzt. Zunächst wird anhand der International Standard Serial Number (ISSN) und des Publikationsjahrs (oder alternativ des Jahrgangs) abgeglichen, ob die vom Verlag gelieferten Datensets in einen Lizenzzeitraum mit entsprechend verhandelter Open-Access-Komponente fallen. Hierzu wurden zu Beginn des Projekts die entsprechenden Titellisten und Jahrgänge manuell aus der EZB-Datenbank extrahiert, im späteren Projektverlauf wurde die EZB prototypisch direkt via Schnittstelle angebunden.

Fallen ISSN- und Jahresprüfung positiv aus, werden anschließend die laut Lizenzvertrag berechtigten wissenschaftlichen Einrichtungen ermittelt. Jede Einrichtung kann mögliche Namensvariationen ihrer Affiliation in ihrem Konto in der so genannten match-config-Datei<sup>17</sup> pflegen. Die von den Verlagen mitgelieferten Affiliationsangaben in den Artikelmetadaten oder, wenn nicht vorhanden, die E-Mail-Domains der korrespondierenden Autorinnen und Autoren werden technisch extrahiert und mit dieser Datei im Konto der Einrichtung abgeglichen. Jede so ermittelte Zuordnung und anschließende Zustellung an das zur Veröffentlichung vorgesehene Repository wird für Verlage und Einrichtungen gleichermaßen protokolliert.

Auf Wunsch der Verlage sollte entgegen der ursprünglichen Antragsidee kein nationales Repository oder „Dark Archive“ eingerichtet werden, sondern das System ähnlich zum Publications Router von Jisc als reines „Push-Forward-System“ designet werden: Somit schickt das System nun unmittelbar Benachrichtigungen über eingegangene Artikel, die dann im nächsten Schritt automatisiert abgeholt werden können, sofern entsprechende (z.B. SWORD-)Schnittstellen konfiguriert sind. Die Metadaten

---

13 Für den letzten Stand der DeepGreen Software vgl. OA-DeepGreen, GitHub, <<https://github.com/OA-DeepGreen>>, Stand: 31.01.2019.

14 Vgl. OPUS 4 – Repository Software, KOBV, <<https://www.kobv.de/entwicklung/software/opus-4/>>, Stand: 31.01.2019.

15 Vgl. DSPACE, DURASPACE, <<https://duraspace.org/dspace/>>, Stand: 31.01.2019.

16 Vgl. OPUS Import-XML, OPUS 4, <<http://www.opus-repository.org/devdoc/import/importxml.html>>, Stand: 31.01.2019.

17 Die match-config-Datei ist eine CSV (Comma-separated values)-Datei, in der die Kriterien hinterlegt werden, mit denen der Artikel einer Einrichtung zugeordnet wird, konkret etwa die verschiedenen Schreibweisen der Affiliation, die E-Mail-Domains, die ORCIDs etc.

und Volltexte der Verlage werden nicht langfristig vorgehalten, sondern wie oben beschrieben einfach weitergeleitet.

Die jetzige Version von DeepGreen ist auf Basis des Jisc Publications Event Router V1, 2016<sup>18</sup> programmiert und läuft unter einem Python Framework Flask v0.9<sup>19</sup>. Die interne Datenverarbeitung wird über Elasticsearch<sup>20</sup> unterstützt. Die Datenannahme ist über sFTP<sup>21</sup> und SWORD v2 in den Metadatenformaten NISO JATS<sup>22</sup> und der Dokumenttypdefinition der Royal Society of Chemistry - DTD RSC möglich. Die Datenlieferung erfolgt über OAI-PMH<sup>23</sup> und SWORD v2 in den Metadatenformaten NISO JATS, RIOXX<sup>24</sup>, Dublin Core<sup>25</sup>, METS/MODS<sup>26</sup>, DTD RSC, OPUS-XML oder eSciDoc-XML<sup>27</sup>. Die Softwareentwicklung ist auf Github unter einer offenen Apache v2.0 Lizenz dokumentiert: <https://github.com/OA-DeepGreen>.

Die technische Anbindung der Verlage funktioniert seit Oktober 2016, die Anbindung der Testrepositorien (auf Softwarebasis von OPUS 4, DSpace und ESciDoc<sup>28</sup>, mit entsprechenden Metadatenkonvertierungen auf das jeweilige Importformat) seit Mai 2017. In Vorbereitung für einen Testbetrieb wurden zudem Standard-Konten für alle Repositorienbetreiber in Deutschland eingerichtet, die seit 2011 an einer der Allianz-Lizenzen mit den Projektpilotverlagen S. Karger und SAGE Publications beteiligt waren bzw. sind (circa 250 wissenschaftliche Einrichtungen).

18 Für die neueste Version der Jisc Publications Event Router Software vgl. JiscPER/jper, GitHub, 09.03.2016, <<https://github.com/JiscPER/jper>>, Stand: 31.01.2019.

19 Flask ist ein Python geschriebenes Webframework, für die neueste Version vgl. Ronacher, Armin: Flask, web development, one drop at a time, flask.pocoo, <<http://flask.pocoo.org/>>, Stand: 31.01.2019.

20 Elasticsearch bezeichnet ein Suchmaschinenprogramm, das frei nachnutzbar zur Verfügung steht, vgl. Elasticsearch, Elasticsearch B.V., <<https://www.elastic.co/de/products/elasticsearch>>, Stand: 31.01.2019.

21 Das Client-Server-Verbindungsprotokoll SSH Secure File Transfer Protocol – sFTP ermöglicht die verschlüsselte Übertragung von Dateien, vgl. SFTP – SSH Secure File Transfer Protocol, ssh.com., <<https://www.ssh.com/ssh/sftp/>>, Stand: 31.01.2019.

22 Die Journal Article Tag Suite – JATS ist ein XML-Metadatenformat für die Beschreibung und den Transfer von Zeitschriftenartikeln, vgl. Journal Article Tag Suite, NCBI, <<https://jats.nlm.nih.gov/>>, Stand: 31.01.2019.

23 Das Open Archives Initiative Protocol for Metadata Harvesting – OAI-PMH ermöglicht den Austausch von Publikationen und deren bibliografischer Metadaten zwischen Online-Repositorien, vgl. Open Archives Initiative Protocol for Metadata Harvesting, Open Archives, <<https://www.openarchives.org/pmh/>>, Stand: 31.01.2019.

24 Vgl. Walk, Paul: the RIOXX metadata application profile and guidelines, rioxx, <<http://www.riox.net/>>, Stand: 31.01.2019.

25 Die Dublin Core Metadata Initiative liefert mit Dublin Core ein standardisiertes Vokabular für die Erzeugung von Metadaten für verschiedenste Dokumentarten, vgl. Dublin Core Metadata Initiative. DCMI Specifications, dublicore, <<http://dublincore.org/specifications/>>, Stand: 31.01.2019.

26 Der Metadata Encoding and Transmission Standard (METS) und das Metadata Object Description Schema (MODS), beide entwickelt von der U.S.-amerikanischen Library of Congress, sind weitere Standards für bibliografische Metadaten, vgl. Cundiff, Morgan; Trail, Nate: Using <METS> and <MODS> to Create XML Standards-based Digital Library Applications, Library of Congress, 18.10.2010, <<https://www.loc.gov/standards/mods/presentations/mets-mods-morgan-ala07/>>, Stand: 31.01.2019.

27 ESciDoc-XML ist das XML-Importformat der Repositoriensoftware eSciDoc, vgl. eSciDoc XML Import, MPG.PuRe-Blog, 31.05.2017, <<http://blog.pure.mpg.de/escidoc-xml-import/>>, Stand: 31.01.2019.

28 Vgl. eSciDoc - The Open Source e-Research Environment, eSciDoc, <<https://www.escidoc.org/>>, Stand: 31.01.2019.

## 5. Zusammenfassung der bisherigen Projektmeilensteine

In der ersten Laufzeit von 2016-2017 wurden bis auf die nachfolgend noch genannte Ausnahme der groß angelegten quantitativen Tests alle gesetzten Ziele der Projektphase erreicht. Das Projektkonsortium evaluierte die technischen und formalen Möglichkeiten der vierzehn Allianz-Lizenzverlage, entwickelte ein Metadatenschema, diskutierte auf Workshops und Fachtagungen die Projektideen mit Expertinnen und Experten aus Bibliotheken, Verlagen, Wissenschaft und der Open Access Community und erreichte das von der DFG gesetzte Ziel einer ersten prototypisch funktionierenden Version der Datendrehscheibe innerhalb eines Jahres.

Qualitative Aussagen zur Funktionsfähigkeit der Drehscheibe konnten durch die erhaltenen Testdaten aus Verlagshand und die drei angebotenen Testrepositorien zum Ende der ersten Projektphase bereits getroffen werden, quantitative jedoch nicht in Gänze. Für einen großangelegten Test der DeepGreen-Funktionalitäten fehlten in der ersten Projektphase noch kritische Datenmengen seitens der Allianz-Lizenzverlage (die mittlerweile retrospektiv für den Zeitraum 2011 bis 2018 für die abgeschlossenen Allianz-Lizenzen von S. Karger und SAGE Publications vorliegen) mit einer relevanten Menge an real angebotenen Repositorien. Dies ist für die zweite Projektphase (2018-2020) geplant, konkret startet DeepGreen mit 50 ausgesuchten Repositorien im Sommer 2019 in eine, für die teilnehmenden Institutionen kostenfreie, erweiterte Testphase.

Ein weiteres Problem ist die Verlässlichkeit der Kennzahlen. So wurden 2017 exemplarisch durch die Abfrage verschiedener Datenbanken zunächst für das Publikationsjahr 2015 alle zugehörigen Artikelzahlen für die FAU, die TU Berlin und das GFZ für die Allianz-Lizenzen seitens S. Karger und SAGE Publications erhoben. Diese Zahlen stellen die Menge der zu erwartenden Artikel dar. Anschließend wurden die Datenlieferungen der beiden Verlage in die DeepGreen-Datendrehscheibe eingespielt und den Repositorien der Projektteilnehmer zugeordnet.

Ein Vergleich zwischen der Anzahl der erwarteten Artikel mit der tatsächlichen Anzahl der von S. Karger und SAGE Publications über die Datendrehscheibe zugeordneten Artikel ergab eine Diskrepanz, die sich auf folgende Ursachen zurückführen lässt: Einige der fehlenden Artikel waren Abstracts in Tagungsbänden ohne DOI und bei einigen Artikeln wich das Publikationsdatum der Onlineveröffentlichung (2015) vom Datum der Printveröffentlichung (2016) ab. Bei einem Artikel enthielten die Metadaten keinerlei Affiliationsangaben und bei zwei Artikeln wurden in der Affiliation Namensvariationen mitgeliefert, die in der match-config-Datei der berechtigten Institution fehlten. Die restlichen Artikel wurden wie erwartet durch die DeepGreen-Datendrehscheibe fehlerfrei zugestellt.

Daraus ergaben sich für DeepGreen als Ergebnis des Testdurchlaufs der ersten Projektphase eine Reihe von Schlussfolgerungen: Erstens können nur Publikationen erfolgreich zugeordnet werden, die eine DOI und entsprechende Affiliationsangaben in den Metadaten besitzen. Zweitens bestimmt der Verlag als grundlegendes Veröffentlichungsdatum einer Publikation das Datum der Printveröffentlichung, was das Ende der Embargofrist zur Veröffentlichung in den Repositorien beeinflusst. Und drittens muss eine an DeepGreen teilnehmende Institution ihre match-config-Datei möglichst



umfangreich pflegen, um eine zuverlässige Zuordnung der Artikel durch die DeepGreen-Datendrehscheibe zu gewährleisten.

## 6. Ausblick auf die zweite Projektphase

Die Planungen für die zweite Förderphase umfassen gemäß dem bewilligten DFG-Folgeantrag

- die Etablierung eines bundesweit agierenden und nachhaltigen Betriebs- und Geschäftsmodells für DeepGreen,
- eine Ausweitung auf andere Lizenzmodelle,
- die Ausweitung auf andere Datenabnehmer, z.B. Fachrepositorien,
- die Optimierung, Konsolidierung und Weiterentwicklung von Workflows sowie der technischen Datendrehscheibe selbst und
- die begleitende Projektkommunikation und Integration aller beteiligten Einrichtungen, Repositorien und Verlage.<sup>29</sup>

An oberster Stelle steht das benannte Ziel, für DeepGreen ein nachhaltiges Geschäfts- und Betriebsmodell zu entwickeln und das Angebot als bundesweiten Dienst (bezogen auf die Allianz-Lizenzen) zu implementieren. Dazu müssen die Beziehungen zwischen Verlagen, Repositorienbetreibern und den Betreibern von DeepGreen rechtlich ausgestaltet werden und es soll eine relevante Zahl von Verlagen mit regelmäßigen Datenlieferungen für den Routinebetrieb gewonnen werden. Hierfür muss auch die technische Infrastruktur konsolidiert und erweitert werden, z.B. hinsichtlich möglicher alternativer Liefer Routinen der Verlage. Auch die Unterstützung weiterer Repositorien-Softwaretypen (wie EPrints<sup>30</sup>, MyCoRe<sup>31</sup>, Fedora<sup>32</sup>, Invenio<sup>33</sup> etc.) ist nötig.

Darüber hinaus soll das Potenzial von DeepGreen mit Blick auf neue, andere Lizenzkontexte eruiert werden, konkret:

- Konsortiallizenzen ohne explizite grüne Open-Access-Komponente,
- FID (Fachinformationsdienste für die Wissenschaft)-Lizenzen,
- goldene Open-Access-Artikel aus reinen Open-Access-Zeitschriften (äquivalent zu den mit Jisc kooperierenden reinen Open-Access-Verlagen),
- Transformationsverträge, Offsetting-Verträge (z.B. DEAL<sup>34</sup>) etc.

Ein dritter Schwerpunkt ist die Frage, ob DeepGreen auch für neue Datenabnehmer geeignet ist, speziell fällt hier der Blick auf Fachrepositorien und Forschungsinformationssysteme. Dazu muss

---

29 Vgl. Koch; Ceynowa; Söllner u. a.: DeepGreen – Open Access Transformation, 2018.

30 Vgl. eprints repository software, EPrints, <<http://files.eprints.org/>>, Stand: 31.01.2019.

31 Vgl. MyCoRe – my content repository, MyCoRe, <<http://www.mycore.de/>>, Stand: 31.01.2019.

32 Vgl. Fedora, DURASPACE, <<https://duraspace.org/fedora/>>, Stand: 31.01.2019.

33 Vgl. Invenio, invenio-software, <<https://invenio-software.org/>>, Stand: 31.01.2019.

34 Vgl. Projekt DEAL, Bundesweite Lizenzierung von Angeboten großer Wissenschaftsverlage, projekt-deal, <<https://www.projekt-deal.de/>>, Stand: 31.01.2019.

ermittelt werden, anhand welcher Datenelemente und Verfahren eine fachliche Identifikation der Artikel möglich ist. Außerdem müssen rechtliche Zusammenhänge geklärt werden und es ist eine Aufwandsanalyse und zugehörige Planung für notwendige technische Spezifikationen, Schnittstellen und Workflows nötig.

Als Partner für die Fachrepositorien haben bis dato das DIPF – Leibniz-Institut für Bildungsforschung und Bildungsinformation Frankfurt am Main mit ihrem Repositorium peDOCS<sup>35</sup>, die Staatsbibliothek zu Berlin mit <intR><sup>2</sup>Dok<sup>36</sup> und das ZBW – Leibniz-Informationszentrum Wirtschaft Kiel mit Econstor<sup>37</sup> Kooperationswünsche signalisiert; für Forschungsinformationsdienste das GESIS – Leibniz-Institut für Sozialwissenschaften Mannheim mit seinem FID SSOAR<sup>38</sup> und die Niedersächsische Staats- und Universitätsbibliothek (SUB) Göttingen mit dem FID GEO<sup>39</sup> sowie für Forschungsinformationsdienste die UB Bielefeld (PUB<sup>40</sup>) und die TIB Hannover (PURE und VIVO<sup>41</sup>). Daneben werden die bestehenden Kooperationen mit der UB Regensburg (Projekt OA-EZB<sup>42</sup>) und mit Jisc (Publications Router<sup>43</sup>) fortgeführt.

Die umfangreiche Ausweitung hat das Ziel, die Aufwände für Repositorienbetreiber sowie Autorinnen und Autoren weiter zu reduzieren, somit insbesondere den grünen Weg von Open Access voranzubringen und, um auf den Titel der Session „Open Access – Zukunftsfähig?“ auf dem Deutschen Bibliothekartag zurückzukommen, dessen Zukunft zu sichern. Ansonsten werden Schattenbibliotheken wie SciHub den rechtlich legitimierte Repositorien den Rang ablaufen.<sup>44</sup>

## Literaturverzeichnis

- 12.181 – Grundsätze für den Erwerb DFG-geförderter überregionaler Lizenzen (Allianz-Lizenzen), DFG, 01.03.2015, <[www.dfg.de/formulare/12\\_181/](http://www.dfg.de/formulare/12_181/)>, Stand: 31.01.2019.
- Becker, Pascal-Nicolas; Bertelmann, Roland; Ceynowa, Klaus u. a.: Questionnaire for effective exchange of bibliographic metadata – current status of publishing houses,

35 Vgl. peDOCS. Open Access Erziehungswissenschaften, pedocs, <<https://www.pedocs.de/>>, Stand: 31.01.2019.

36 Vgl. <intR><sup>2</sup> [S]. Fachinformationsdienst für internationale und interdisziplinäre Rechtsforschung, <intR><sup>2</sup> [S], <<https://intr2dok.vifa-recht.de/content/index.xml?jsessionid=44840C210AB98A29DFE1D1E60842C6C7>>, Stand: 31.01.2019.

37 Vgl. ECONSTOR, econstor, <<https://www.econstor.eu/?locale=de/>>, Stand: 31.01.2019.

38 Vgl. Social Science Open Access Repository, gesis, <<https://www.gesis.org/ssoar/home/>>, Stand: 31.01.2019.

39 Vgl. Fachinformationsdienst Geowissenschaften der festen Erde, FidGeo, <<https://www.fidgeo.de/>>, Stand: 31.01.2019.

40 Vgl. PUB - Publikationen an der Universität Bielefeld, Universität Bielefeld, <<https://pub.uni-bielefeld.de/>>, Stand: 31.01.2019.

41 Vgl. VIVO, TIB Hannover, <<https://vivo.tib.eu/vivo/>>, Stand: 31.01.2019.

42 Vgl. OA-EZB: Open-Access-Services der Elektronischen Zeitschriftenbibliothek (DFG-Projekt), Universität Regensburg, <<https://www.uni-regensburg.de/bibliothek/projekte/oa-ezb/>>, Stand: 31.01.2019.

43 Vgl. Publications Router. Helping institutions capture their research articles onto their open repositories, Jisc, <<https://pubrouter.jisc.ac.uk/>>, Stand: 31.01.2019.

44 Vgl. Green, Toby: We've failed: Pirate black open access is trumping green and gold and we must change our approach, in: Learned Publishing 30 (4), 2017, S. 325-329, Online: <<https://doi.org/10.1002/leap.1116>>, hier S. 326.

Dokumentenserver des Zuse Institute Berlin (ZIB), 29.09.2016, <<http://nbn-resolving.de/urn:nbn:de:0297-zib-60419>>.

- Der freie Zugang zu wissenschaftlicher Information. Positionen, open-access.net, 31.01.2019, <<https://open-access.net/informationen-zu-open-access/positionen/>>, Stand: 31.01.2019.
- Green, Toby: We've failed: Pirate black open access is trumping green and gold and we must change our approach, in: Learned Publishing 30 (4), 2017, S. 325-329, Online: <<https://doi.org/10.1002/leap.1116>>.
- Herb, Ulrich: Open Access – Erfolge und nicht-intendierte Folgen, BIB OPUS Publikationsserver, 04.06.2018, <<https://nbn-resolving.org/urn:nbn:de:0290-opus4-36015>>.
- Keuenhof, Maximilian: Open Access in der Kritik, bibliotheksnews, 14.06.2018, <<https://bibliotheksnews.com/2018/06/14/open-access-in-der-kritik/>>, Stand: 31.01.2019.
- Oberländer, Anja: Förderung von Open Access über institutionelle Infrastrukturen, insbesondere Repositorien, in: Söllner, Konstanze; Mittermaier, Bernhard (Hg.): Praxishandbuch Open Access (E-Book), Berlin; Boston 2017, S. 137-145, <<https://doi.org/10.1515/9783110494068-016>>.
- Schäffler, Hildegard: Open Access in konsortialer Perspektive, in: Söllner, Konstanze; Mittermaier, Bernhard (Hg.): Praxishandbuch Open Access (E-Book), Berlin; Boston 2017, S. 197-204, <<https://doi.org/10.1515/9783110494068-023>>.
- Tobias, Regine: Die Quote kommt – Einwerbung von Open-Access-Publikationen durch nutzernehe Workflows im Repository, BIB OPUS Publikationsserver, 04.06.2018, <<https://nbn-resolving.org/urn:nbn:de:0290-opus4-35961>>.