

Birgit Schlegel

# Automatisierter Import von geförderten Open-Access-Artikeln in Repositorien – eine Herausforderung für Repositoriums-Betreiber\*innen?

## Automated Import of Funded Open-Access Papers into Repositories – A Challenge for Repository Operators?

<https://doi.org/10.1515/bd-2022-0100>

**Zusammenfassung:** Der Beitrag beschreibt die Schwierigkeiten bei der Umsetzung einer automatisierten Übernahme aller über Publikationsgelder finanzierten Open-Access-Zeitschriftenartikel in institutionelle Repositorien. Sind Verlags-Dashboards, die den wissenschaftlichen Einrichtungen im Rahmen von Vertragsvereinbarungen für die Bearbeitung der eingehenden Verifizierungsanfragen zur Verfügung gestellt werden, in diesem Zusammenhang hilfreich?

**Schlüsselwörter:** Institutionelles Repositorium, Open Access, Publikationsgebühren, Verlagsverträge, Schnittstellen, Datenimport, Metadaten, Verlagsdashboards



**Abstract:** The article discusses the complexities of implementing the automatic import of Open Access journal articles with a publication grant into institutional repositories, and asks whether publisher dashboards provided to academic institutions as part of contract agreements on the management of incoming verification requests are a useful tool to deal with the issue.

**Keywords:** Institutional repository, open access, publication fee, publishing contracts, interface, data import, metadata, publisher dashboards

Um Wissenschaftler\*innen eine große Bandbreite an Möglichkeiten des kostenfreien Open-Access-Publizierens in wissenschaftlichen Zeitschriften zu bieten, schließen Bibliotheken und wissenschaftliche Einrichtungen Verlagsvereinba-

---

**Birgit Schlegel:** [schlegel@ub.fu-berlin.de](mailto:schlegel@ub.fu-berlin.de)

 Open Access. © 2022 bei den Autoren, publiziert von De Gruyter. 

Dieses Werk ist lizenziert unter der Creative Commons Namensnennung 4.0 International Lizenz.

rungen<sup>1</sup>, institutionelle Mitgliedschaften<sup>2</sup> oder konsortiale Verträge mit wissenschaftlichen Verlagen wie de Gruyter, Cambridge University Press, Frontiers, Copernicus und anderen.

Vertraglich festgelegt werden neben den allgemeinen Geschäftsbedingungen die Zahlungsmodalitäten (prepayment, direct invoicing, postpayment), mögliche Rabattierungen und Informationen über die zur Verfügung stehenden Zeitschriftenkontingente, in denen die in der jeweiligen Einrichtung affilierten Autor\*innen kostenfrei publizieren können.

Worüber grundsätzlich bei Vertragsabschluss nicht gesprochen wird, sind mögliche Schnittstellen für eine automatisierte Einspielung der finanzierten und publizierten Artikel in die Publikationsserver der teilnehmenden Einrichtungen. Technische Erfordernisse wie Schnittstellen für eine automatisierte Einspielung der finanzierten und publizierten Artikel in die Publikationsserver der teilnehmenden Einrichtungen sind in der Regel nicht Vertragsbestandteil, sie werden erst nach Vertragsbeginn eventuell mit dem technischen Support der Verlage erörtert.

Erst nachdem die geförderten Artikel als Verlagsversion erschienen sind, wird seitens der Einrichtung der Ruf nach einer möglichst mit wenig Arbeit verbundenen Übernahme der Artikel in das hauseigene institutionelle Repositorium laut.

Gemäß dem Zweitveröffentlichungsrecht § 38 UrhG Abs. 4 sind Autor\*innen<sup>3</sup> aus teilnehmenden Einrichtungen ohne Mehrkosten berechtigt, ihre in den lizenzierten Zeitschriften publizierten Artikel in der durch den Verlag publizierten Form sofort nach Erscheinen in ein institutionelles Repositorium ihrer Wahl einzupflegen und kostenlos zugänglich zu machen.

Stellen Autor\*innen einen Antrag auf finanzielle Förderung eines Open-Access-Artikels, so kann die Einrichtung die Wissenschaftler\*innen auf die Möglichkeit der selbständigen Anmeldung der Metadaten gezielt aufmerksam machen. Die Universitätsbibliothek der Freien Universität Berlin hat diesen Hinweis bspw. in ihren Antragsformularen<sup>4</sup> als Pflichtfeld mit folgendem Wortlaut hinterlegt: „Der Artikel wird von mir unverzüglich über Refubium zur Verfügung gestellt.“ Ob die Autor\*innen dieser Bitte letztendlich nachkommen, wird im Laufe des Textes erörtert.

---

1 [https://www.fu-berlin.de/sites/open\\_access/finanzierung/verlagsvereinbarungen/index.html](https://www.fu-berlin.de/sites/open_access/finanzierung/verlagsvereinbarungen/index.html) [Zugriff: 11.08.2022].

2 [https://www.fu-berlin.de/sites/open\\_access/finanzierung/institutionelle\\_mitgliedschaften/index.html](https://www.fu-berlin.de/sites/open_access/finanzierung/institutionelle_mitgliedschaften/index.html) [Zugriff: 11.08.2022].

3 <https://open-access.network/informieren/publizieren/open-access-publizieren> [Zugriff: 11.08.2022].

4 [https://www.fu-berlin.de/sites/open\\_access/finanzierung/publikationsfonds/Formular\\_antrag\\_neu\\_0320.pdf](https://www.fu-berlin.de/sites/open_access/finanzierung/publikationsfonds/Formular_antrag_neu_0320.pdf) [Zugriff: 11.08.2022].

Mit der Frage einer vollautomatisierten Einspielung in das institutionelle Repositorium, dem Refubium<sup>5</sup>, hat sich die Universitätsbibliothek der Freien Universität Berlin bereits in 2012 seit Beginn der Förderung von Open-Access-Publikationen beschäftigt. Um möglichst wenig manuelle Vor- und Nacharbeit zu investieren, sollten die Metadaten und die PDF-Dateien der Verlagsversionen in das Repositorium eingespielt werden. Dies geschah und geschieht über einen simplen Datenimport. Anfangs gestaltete sich diese Übertragung etwas holprig, da der Bibliothek ein Zugang zu einem FTP-Server zur Verfügung gestellt wurde, bei dem kontinuierlich manuell angefragt werden musste. Zudem fehlten seitens der Verlage Ansprechpartner\*innen, so dass kein kontinuierlicher Workflow aufgebaut werden konnte.

In Zusammenarbeit mit Verlagen wie Copernicus<sup>6</sup> und MDPI<sup>7</sup> wurde eine erweiterte Möglichkeit des simplen Datenimports gefunden. Copernicus legt via SFTP<sup>8</sup> die Dokumente im PDF-Format und die dazugehörigen Metadaten im XML<sup>9</sup>-Format ab. Über ein täglich auszuführendes Skript, einen sogenannten Cronjob, wird das XML-Datenformat mithilfe von XSLT<sup>10</sup> vom MODS<sup>11</sup>-Format in das DSpace-interne Metadatenformat (DIM<sup>12</sup>) umgewandelt und danach in den DSpace Workflow des Refubiums importiert.

MDPI liefert die Dokumente ebenfalls im PDF-Format und die Metadaten als XML via SWORD Schnittstelle<sup>13</sup>. Zuerst wird das XML vom METS/EPDCX<sup>14</sup>-Format in das DSpace-interne DIM-Format mit XSLT gewandelt und dann ebenso ins Refubium importiert. Bevor die Daten im Workflow sichtbar sind, nimmt die IT-Abteilung des Refubiums ggf. noch einmal eine Kuratierung der Metadaten vor.

Obgleich wir in beiden Fällen nicht von Schnittstellen sprechen, gestaltet sich die Arbeit der Editor\*innen durch diese Art des simplen Datenimports wesentlich effizienter als die reine manuelle Eingabe der Metadaten und PDF-Dateien über die Eingabeformulare des Repositoriums.

---

5 <https://refubium.fu-berlin.de> [Zugriff: 11.08.2022].

6 <https://publications.copernicus.org> [Zugriff: 18.08.2022].

7 <https://www.mdpi.com> [Zugriff: 18.08.2022].

8 Secure File Transfer Protocol.

9 Extensible Markup Language dient als text-basiertes Format für den Austausch strukturierter Information.

10 XSLT dient zur Generierung von HTML-Code aus den XML-Daten.

11 Metadata Object Description Schema (*MODS*) ist ein XML-Format für bibliografische Metadaten.

12 Dspace internal metadata.

13 Simple Webservice Offering Repository Deposit.

14 XML-Format zur strukturierten Beschreibung von digitalen Objekten, EPDCX oder EPrints DC XML wird als voreingestelltes Metadaten-Format empfohlen.

Das Editor\*innenteam der Bibliothek ergänzt in beiden Fällen nur einige wenige finanzierungsrelevante lokale Metadaten und gibt die geförderten Verlagsartikel anschließend frei. Eine interne Workflowdokumentation, die die Abläufe vom Zeitpunkt der Einreichung des Artikels über die Bezahlung bis hin zur Verfügbarmachung im Refubium der Freien Universität Berlin beschreibt, wird in regelmäßigen Abständen aktualisiert. Die zusätzliche Erstellung und die Pflege von Excel-Listen entfällt – demnach findet hier ein sehr effizienter Workaround statt.

Gleichermaßen ist es möglich, einen Datenimport unter Zuhilfenahme von csv-Dateien der Metadaten zu vollziehen. Anhand einer Ausfüllhilfe für Editor\*innen als Abgleich der Metadaten werden die Daten aufbereitet und anschließend in das Repositorium importiert.

Welche Alternativen gibt es neben der Möglichkeit des Datenimports?

1. Die Wissenschaftler\*innen informieren ihre Einrichtung nach Abschluss der Förderung über den Erscheinungstermin ihres Artikels. Damit sie dieser Aufgabe nachkommen, kann diese Möglichkeit zum Zeitpunkt der Antragstellung im Webformular<sup>15</sup> als anklickbares Pflichtfeld verankert werden, wie auch am Beispiel der FU Berlin bereits erwähnt. Ein Nachteil dabei ist, dass die Autor\*innen zum Zeitpunkt des Erscheinens ihrer geförderten Artikel teils gar nicht mehr an der Einrichtung tätig sind oder sich aus zeitlichen Gründen nicht mit dieser Aufgabe befassen möchten. In der Tat meldet nur ein sehr geringer Teil der Autor\*innen an der FU Berlin ihre durch die UB geförderten Artikel selbständig im institutionellen Repositorium an.
2. Die Bibliothek setzt sich nach erfolgreicher Abwicklung des Bezahlvorgangs eine Erinnerung entweder per E-Mail an eine Funktionsadresse oder via Ticketsystem. So wird sichergestellt, dass in regelmäßigen Zeiträumen von dem Editor\*innenteam überprüft wird, ob der betreffende Artikel bereits erschienen ist. Anmeldung der Metadaten und Hochladen der Verlagsversion übernimmt das Editor\*innenteam der Bibliothek, das durch passgenau aufeinander abgestimmte Workflows dabei unterstützt wird.
3. Die Bibliothek ist Vertragspartner\*in von Deep Green<sup>16</sup> und nutzt die Sword-Schnittstelle der teilnehmenden Verlage. Deep Green bietet als Datenlieferdienst für Projektpartner\*innen Wege zur automatisierten Übernahme in das hauseigene Repositorium. Die Datenlieferungen im Deep Green-Portal können von der jeweiligen am Projekt teilnehmenden Einrichtung selbst definiert werden. Dies geschieht über einen Verlagsfilter im Deep Green-Portal.

---

<sup>15</sup> [https://www.fu-berlin.de/sites/open\\_access/finanzierung/publikationsfonds/Formularantrag\\_neu\\_0320.pdf](https://www.fu-berlin.de/sites/open_access/finanzierung/publikationsfonds/Formularantrag_neu_0320.pdf) [Zugriff: 11.08.2022].

<sup>16</sup> <https://info.oa-deepgreen.de> [Zugriff: 11.08.2022].

In welcher Form die Publikationen in das Repositorium gelangen, bleibt den Einrichtungen überlassen. Zum einen besteht die Möglichkeit der sofortigen Freigabe im Produktivsystem, zum anderen ist es vorteilhafter, die einzelnen Metadaten in den Workflow zu legen, um sie vor ihrer Freigabe im Produktivsystem des Servers einer finalen formalen Prüfung zu unterziehen. Bei der geschilderten Datenübertragung handelt es sich um konfigurierte Schnittstellen. Allerdings sind diverse manuelle Korrekturen der Metadaten, die sich z. B. auf die Ergänzung finanzierungsrelevanter Felder beziehen, sowie eine zusätzliche Ausfilterung bestimmter Dokumentarten vor der Freisaltung nötig.

4. Die Bibliothek nutzt die Bibliothekssoftware Alma für alle rechnungsrelevanten und etatbezogenen Vorgänge bei Open-Access-Artikeln. Auch bei dieser Art der Abwicklung wird nicht alles abgedeckt, weiterhin ist das Führen von Excel-Listen notwendig.

Da derzeit keine Gesamtlösung existiert, stellte sich die Frage, ob eventuell Dashboards der Verlage für eine automatisierte Datenübertragung der fertigen Verlagsversionen in Repositorien hilfreich sein könnten.

Dashboards bieten eine gute Übersicht hinsichtlich der Verifizierung von Artikeln. Sie unterscheiden sich in ihren Funktionen allerdings zum Teil stark voneinander. Allen gemeinsam ist, dass Änderungen an den bestehenden Detailinformationen, nachträgliche Änderungen bspw. fehlerhafter Institutionsangaben in der Regel nicht umsetzbar sind. Sonstige Änderungen oder Ergänzungen in den bestehenden Tools gestalten sich schwierig und sind je nach Vertragswerk ein längerer Abstimmungsprozess zwischen der jeweiligen Einrichtung, den Konsortialpartner\*innen und dem Verlag.

Exemplarisch werden die Verlagsdashboards von Copernicus, Wiley und das CCC-Institution-Portal von de Gruyter und Cambridge kurz vorgestellt.

Während Plos One<sup>17</sup> mit seinem Verlagsdashboard nur eine einfache Verifizierung erlaubt, ist die Arbeit mit dem MDPI-Tool<sup>18</sup> wesentlich komfortabler. Das MDPI-Tool bietet eine Artikelübersicht (Overview Articles), die eine Filterung nach den verschiedensten Kriterien zulässt. In der zu verifizierenden Artikelübersicht können seitens der Editor\*innen Notizen gesetzt werden, die Autor\*innen können kontaktiert werden und es besteht die Möglichkeit, alle Rechnungsdetails tagesaktuell einzusehen. Jeder Einrichtung steht eine Art Reglersystem zur Verfügung. Steht der Regler auf grau, erhält der corresponding author die Rechnung

---

17 <https://app.sigmacomputing.com/plos> [Zugriff: 11.08.2022].

18 <https://susy.mdpi.com> [Zugriff: 11.08.2022].

zugeschickt. Sobald Wissenschaftler\*innen während des Submission-Prozesses ihre Einrichtung anwählen und damit signalisieren, für die zentrale Rechnungslegung zugelassen zu sein, wird der Radio-Button bei der Akzeptanz des Artikels auf grün gesetzt. Sollte ein Artikel fälschlicherweise auf „grün“ gesetzt worden sein, reicht zum Rückgängigmachen eine kurze Mail an MDPI.

Das CC Rightslink, das Management-Tool<sup>19</sup>, das von CUP<sup>20</sup> und de Gruyter genutzt wird, erlaubt eine Verifizierung der eingereichten Artikel inklusive einer Kostenansicht sowie eine Suchfunktion, die die Auswahl bestimmter Reports erlaubt. Diese Reports werden innerhalb weniger Minuten erstellt. Das Tool bietet keine Möglichkeit zur Übernahme von publizierten Artikeln, auch eine Information über das Erscheinen der Artikel wird nicht präsentiert.

Das Copernicus-Dashboard<sup>21</sup> bietet detaillierte Informationen zum Manuskript Overview und weiteren Informationen. Verifizierungsangaben sind an der FU Berlin an ein Ticketsystem gekoppelt.

Das seit 2021 existierende Dashboard von Frontiers<sup>22</sup> zeigt alle zu verifizierenden Artikel. Weitere Möglichkeiten der Bearbeitung bestehen hier nicht.

Das Verlagsdashboard von Wiley<sup>23</sup> stellt den Einrichtungen für die beiden separaten Gold- und Hybrid-Accounts Autor\*innenangaben und Publikationsdaten von Artikeln zur Verfügung. Anhand derer wird vom Open-Access-Team der Bibliothek überprüft, ob ein Artikel zur Open-Access-Publikation berechtigt und damit förderfähig ist. Die eingehenden Verifizierungsanfragen erscheinen im Dashboard im jeweiligen Account und werden bspw. an der Universitätsbibliothek der FU Berlin über ein Ticketsystem kommuniziert. Die Anfrage erscheint im Dashboard als noch nicht beantwortet, request als „pending“ markiert.

Die Aufgabe der Einrichtung ist es, alle Eingänge zeitnah zu prüfen, um sie entweder zu genehmigen (approve) oder abzulehnen (deny).

Dashboards sind also lediglich für einen kleinen Teil im Bearbeitungsvorgang der finanzierenden Einrichtung hilfreich. Ein konsequentes Bearbeiten von der Verifizierung eines Artikels bis hin zur Bezahlung und der Übermittlung der erschienenen Verlagsversion via Schnittstelle aus dem Dashboard heraus ist nicht möglich, zumal Dashboards nicht mit anderen Bibliothekssoftware- oder Discoverysystemen wie Alma oder Primo korrespondieren.

---

<sup>19</sup> <https://apcfunding.copyright.com> [Zugriff: 11.08.2022].

<sup>20</sup> <https://www.cambridge.org> [Zugriff: 11.08.2022].

<sup>21</sup> <https://administrator.copernicus.org/authentication.php> [Zugriff: 11.08.2022].

<sup>22</sup> <https://institutional-membership.frontiersin.org/my-funding-requests> [Zugriff: 11.08.2022].

<sup>23</sup> <https://authorservices.wiley.com/index.html#login?returnUrl=https://openaccessaccount.wiley.com> [Zugriff: 11.08.2022].

Muss also demnach weiterhin mit Einzellösungen wie simplen Datenimporten und ergänzenden Excel-Listen gearbeitet werden?

Kommerzielle Anbieter wie Lehmanns Media<sup>24</sup> und Chronos Hub oder Knowledge Unlatched<sup>25</sup> möchten diese Lücke schließen und bieten Komplett-Services für das Management von Open-Access-Publikationen an. Diese Gesamtlösungen berücksichtigen u. a. auch die Bedürfnisse der Autor\*innen, die sich in den seltensten Fällen mit den Bedingungen, Richtlinien des Publizierens beschäftigen wollen. Bibliotheken wird ein Servicepaket vom Verifizieren, Bezahlen bis zur Datenanalyse und der Erstellung von Nutzungsstatistiken angeboten. Der Preis berechnet sich nach dem Volumen.

Wenn damit auch Lösungen zur automatisierten Übernahme der publizierten Artikel in die Repositorien der jeweiligen Kunden Teil des Gesamtkonzepts wären, dann könnte dies ein Lösungsansatz für die anfangs geschilderte Thematik für Repositorien sein. Eine Art Ampelsystem oder Button zur Einspielung der Metadaten der Verlagsversion aus den Verlagstools heraus wäre eine denkbare Lösung, um den Workflow abzurunden.

In jedem Fall hilfreich für die Entwicklung potentieller Prozesse zur vollautomatisierten Übernahme ist eine gute interne Kommunikation zwischen dem Editor\*innenteam des jeweiligen Repositoriums und dem IT-Support der jeweiligen Einrichtung als Voraussetzung für die Betreuung der Datenimporte bzw. Schnittstellen und einer damit verbundenen kontinuierlichen Anpassung aller Updates. Detaillierte Arbeitsbeschreibungen sollten die Grundlage für die Arbeit der Editor\*innen bilden. Als hilfreich erweisen sich parallele Testinstanzen, um Workflows fundiert vor der Produktionsaufnahme auszuprobieren.



**Birgit Schlegel**

Universitätsbibliothek der Freien Universität Berlin

Abt. Dienste für Forschung

Leitung Team Hochschulschriftenstelle u. Dokumentenserver

Garystraße 39

14195 Berlin

Deutschland

E-Mail: [schlegel@ub.fu-berlin.de](mailto:schlegel@ub.fu-berlin.de)

---

24 <https://lehmannsbib.de> [Zugriff: 11.08.2022].

25 <https://knowledgeunlatched.org> [Zugriff: 11.08.2022].