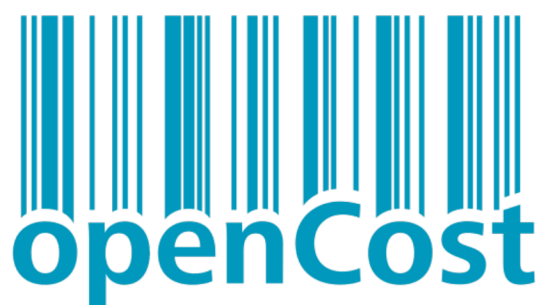


openCost:
Automatisierte, standardisierte Lieferung und offene Bereitstellung
von Publikationskosten und Verlagsvereinbarungen

– Zwischenbericht zum DFG-Projekt Nr. 457354095 –



Funded by

DFG Deutsche
Forschungsgemeinschaft
German Research Foundation

1. Allgemeine Angaben zum Projekt openCost

Titel (Projektnummer): „openCost: Automatisierte, standardisierte Lieferung und offene Bereitstellung von Publikationskosten und Verlagsvereinbarungen“ (Projektnummer: 457354095)

Berichtszeitraum (gesamte Förderdauer): 16.2.2022 - 16.2.2025

Namen der Antragstellenden und Dienstanschriften: Dr. Gernot Deinzer (Universität Regensburg, Universitätsstraße 31, 93053 Regensburg), Dirk Pieper (Universität Bielefeld, Universitätsstraße 25, 33615 Bielefeld), Dr. Alexander Wagner (Deutsches Elektronen-Synchrotron DESY, Standort Hamburg, Notkestraße 85, 22607 Hamburg)

Namen der Mitverantwortlichen: Julia Bartlewski, Christoph Broschinski, Cornelia Lang, Bianca Schweighofer, Colin Sippl, Lisa-Marie Stein, Silke Weisheit

Namen der Kooperationspartnerinnen und -partner: JOIN² (siehe 5.). Das Forschungszentrum Jülich (FZJ) ist als JOIN²-Partner in das Projekt eingebunden. Darüber hinaus besteht mit der Zentralbibliothek des FZJ im Zuge des Publikationskostenmonitorings, mit dem sie im Rahmen des DFG-Programms „Open-Access-Publikationskosten“ (OAPK) beauftragt ist, und dem ebenfalls von Jülich betriebenen Open Access Monitor seit Projektbeginn eine enge Abstimmung.

2. Zusammenfassung

Ziel von openCost ist die Schaffung einer technischen Infrastruktur, mit der Publikationskosten über standardisierte Schnittstellen und Formate frei zugänglich ausgetauscht werden können. Dadurch soll Kostentransparenz auf institutioneller, nationaler und intern. Ebene ermöglicht werden.

Dafür wurde unter Einbeziehung internationaler Expertise ein Metadatenschema erarbeitet, mit dem die Publikationskosten einer wiss. Einrichtung in strukturierter Form erfasst, abgebildet und abgefragt werden können. Das grundlegende OpenAPC-Metadatenschema wurde zum openCost-Metadatenschema erweitert, um auch mit Open-Access-Publikationsgebühren einhergehende Zusatzkosten abbilden zu können. Der Datenaustausch erfolgt über eine OAI-PMH-Schnittstelle.¹ Für openCost fungiert OpenAPC als *Service Provider*, während die Projektpartner*innen und einige Einrichtungen von JOIN² *Data Provider* darstellen. Ein weiterer Projektschwerpunkt ist die Erweiterung der Elektronischen Zeitschriftenbibliothek (EZB) um Funktionen rund um Publikationskosten. Dazu wird die OpenAPC-Schnittstelle in die EZB eingebunden, sodass automatisiert aggregierte Kosteninformationen angezeigt werden können. Zudem wurde die Möglichkeit geschaffen, institutionsspezifische Informationen zur Übernahme von Publikationskosten abzulegen. Diese Informationen sollen einrichtungsspezifisch bei den einzelnen Zeitschriften angezeigt und nachnutzbar angeboten werden. Die EZB kann damit als zentrale Informationsplattform für die Bereitstellung und Vermittlung an die Forschenden verwendet werden.

¹ Siehe <https://www.openarchives.org/OAI/openarchivesprotocol.html>.

Within the scope of openCost we are creating a technical infrastructure that allows us to comprehensively record all costs involved in scientific publishing and subsequently make them freely accessible by means of standardized interfaces. This will enable cost transparency on an institutional, national, and international level.

Therefore, with the support of (inter-)national experts, we have developed a standardized metadata schema to record, retrieve and map all publication costs of a scientific institution in a structured form. The basic OpenAPC metadata schema was extended to the openCost metadata schema to be able to map additional costs associated with Open Access publication fees. The data is exchanged via an OAI-PMH interface. OpenAPC acts as a *service provider* for openCost, while the project partners and several JOIN² institutions are *data providers*. Another focus of the project is the extension of the Electronic Journals Library (EZB) to include functions regarding publication costs. To achieve this, we integrate the OpenAPC interface into the EZB to display summarized cost information. Additionally, we are extending the EZB so that institutions may enter information on publication funding for specific journals as well as institution-specific agreements in the EZB administration interface. Thus, all participating institutions can use the EZB as a central platform for distributing and communicating this OA information to researchers.

3. Arbeits- und Ergebnisbericht

Im Folgenden wird ein Überblick über die Entwicklung, Durchführung sowie Ergebnisse des Projekts geliefert.

3.1. Ausgangslage und Zielsetzung des Projekts

Um den freien und schnellen Zugang zu wissenschaftlichen Ergebnissen sowie publizierten Informationen nachhaltig zu verbessern, befindet sich die wissenschaftliche Publikationslandschaft in einem Transformationsprozess zu Open Access (OA)². Dies geht mit einer Abkehr von subskriptionsbasierten und Hinwendung zu publikationsgebühren- oder einrichtungsfinanzierten Geschäftsmodellen einher und verändert neben den Rollen der Akteur*innen grundlegend Prozesse und Finanzströme. Die aktuelle Umbruchphase ist geprägt von einer Vielzahl an Geschäftsmodellen und Arbeitsabläufen sowie einer häufig aufwändigen und unvollständigen Erfassung der Publikationskosten. Um rechtzeitig Fehlentwicklungen begegnen und die angestrebte Veränderung nachhaltig gestalten zu können, ist frühzeitig umfassende Kostentransparenz notwendig. Die Vielzahl individueller Vorgänge setzt einen hohen Grad an Automatisierung voraus. openCost setzt hier durch die Schaffung eines maschinenlesbaren Austauschformats und entsprechender

² Die Dynamik dieses Wandels wird u. a. deutlich in Forderungen und Empfehlungen intern. Organisationen und Förderprogrammen wie dem Rahmenprogramm für Forschung und Innovation der EU, in dem die vertragliche Verpflichtung zu OA verankert ist. Auch der Wissenschaftsrat hat 2022 „Empfehlungen zur Transformation des wissenschaftlichen Publizierens zu Open Access“ (DOI: 10.57674/fyrc-vb61) verabschiedet. Vor diesem Hintergrund und der Förderlinie „Open Access Publizieren“ etablieren zahlreiche wiss. Einrichtungen Publikationsfonds und Workflows zur zentralen Bearbeitung.

Schnittstellen an. Die so ermöglichte Offenlegung von Kostenstrukturen lässt Vergleiche auf lokaler sowie (inter-)nationaler Ebene zu, macht Preissteigerungen transparent und verbessert die Verhandlungsposition der Bibliotheken und Konsortien.

3.2. Erarbeitung und Vorstellung eines Metadatenschemas

Ziel von openCost ist ein standardisiertes Metadatenschema, das von möglichst vielen Einrichtungen genutzt wird.³ Deshalb wurden zunächst mithilfe der Community Probleme in Bezug auf die Kostenerfassung identifiziert und daraus Bedarfe ermittelt [6]. Die Ergebnisse legten den Grundstein für die Projektarbeit und den Workshop „openCost: the road to publication cost transparency“⁴, zu dem (inter-)nationale Expert*innen eingeladen wurden, um ihre Perspektiven und Bedürfnisse bzgl. Publikationskosten und deren Erfassung einzubringen. Daneben präsentierten die Projektpartner*innen einen ersten Entwurf eines Metadatenschemas für gebührenpflichtige Einzelartikel als Grundlage für den Austausch mit den rund 40 Teilnehmenden. Alle Beiträge stehen auf der Veranstaltungsseite⁵ zur Verfügung, die Ergebnisse wurden als Conference Proceedings [9] veröffentlicht.

Im Februar 2023 veröffentlichte das Projektteam ein Metadatenschema zur Erfassung von Kostendaten bei Zeitschriftenartikeln,⁶ wobei dieses bereits Kostenpositionen abdeckte, die über OA-Gebühren hinausgehen.⁷ Das Schema wurde mit der Community in drei Online-Veranstaltungen diskutiert und anschließend auf Basis der Ergebnisse⁸ angepasst.⁹ Um die (inter-)nationale Fachcommunity intensiv und breit einzubinden, wurde das modifizierte Schema auf Kongressen und Veranstaltungen präsentiert (vgl. 3.6.). Zusätzlich zu den Kosten, die auf Artekebene zugeordnet werden können, sollten auch Kosten aus sog. Transformationsverträgen (TV) abgebildet werden können. Ein entsprechendes Metadatenschema für Verträge, das gleichsam für OA, TV und Subskriptionsverträge anwendbar ist, wurde der Community im August 2023 präsentiert.¹⁰

Aus der Community kam wiederholt die Frage auf, ob das openCost-Schema alternativ zum für OAI-PMH benötigten XML-Format auch im JSON-Format darstellbar sei.¹¹ Dies führte zur attributlosen Neumodellierung des Schemas, die diese Konversionen in eindeutiger Form erlaubt.

³ Das openCost-Format wurde als Austauschformat für Kostendaten konzipiert, wobei Daten so sparsam wie möglich und so umfassend wie nötig dargestellt werden sollen. Dies sichert eine nahtlose Integration in andere Formate.

⁴ Der Expert*innenworkshop fand von 5. bis 7.10.2022 am DESY in Hamburg statt. Es beteiligten sich Sprecher*innen aus acht Ländern (DE: Transform2Open, AUT: AT2OA2, NL: OA Switchboard, UK: JISC, USA: California Digital Library, CAN: OurResearch, NOR: Sikt, FIN: The National Library of Finland).

⁵ Vgl. Veranstaltungswebsite, URL: <https://indico.desy.de/event/35620/overview>.

⁶ Vgl. <https://www.opencost.de/aktuelles/metadataschema-rfc/>.

⁷ Z. B. Zusatzkosten wie page oder color charges, die sowohl in Subskriptions- als auch in OA-Modellen auftreten.

⁸ Vgl. <https://www.opencost.de/aktuelles/ergebnisse-online-diskussionen/>.

⁹ Z. B. terminologische Änderungen (vgl. <https://github.com/opencost-de/opencost/issues/37>) oder die Möglichkeit, sowohl Rechnungs- als auch Bezahldatum anzugeben (vgl. <https://github.com/opencost-de/opencost/issues/44>).

¹⁰ Siehe <https://www.opencost.de/aktuelles/transaktionsvertraege/>. Um Community-Feedback einzuholen organisierte das Projektteam im Oktober 2023 eine weitere Online-Diskussionsrunde, deren Anregungen in das Schema einfließen.

¹¹ Vgl. z. B. <https://github.com/opencost-de/opencost/issues/41>.

3.3. Datentransfer im openCost Format

Die am DFG-Programm „Open-Access-Publikationskosten“ teilnehmenden Einrichtungen lieferten im März 2023 erstmals Daten zu den 2022 geförderten Publikationen an das FZJ mit der Möglichkeit der Datenweitergabe an OpenAPC. Diese erfolgte im August 2023 im openCost-Format. Die Gold-OA-Daten wurden anschließend bis Januar 2024 in OpenAPC verarbeitet. Auch die Übernahme der hybriden DEAL-Daten konnte auf technischer Ebene bereits erfolgreich realisiert werden. Aufgrund des frühen Berichtszeitpunkt und den damit verbundenen Abweichungen zu den von der MPDLS versendeten PABA-Rechnungen, auf deren Basis die Meldung an OpenAPC bislang immer erfolgte, wurde in Absprache mit dem FZJ einmalig ein angepasster Workflow entwickelt, über den alle betroffenen Einrichtungen von OpenAPC informiert wurden. Die so von den Einrichtungen zur Verfügung gestellten aktualisierten hybriden DEAL-Daten wurden bis März 2024 in OpenAPC eingepflegt.¹² Zur Darstellung bzw. Meldung der hybriden DEAL-Daten und insbesondere zum Umgang mit „Nach- und Rückzahlungen“ ist für 2024 ein thematischer Workshop geplant, der von OpenAPC, openCost und dem FZJ ausgerichtet wird.

Das OAPK-Programm erfordert ein umfangreiches Monitoring, welches insbesondere auch die Funder-Perspektive abbildet. Die Aufnahme entsprechender Elemente steht jedoch dem Fokus von openCost auf Kosten sowie einer Einbettung in andere Formate im Wege. Auch müssen diese Anforderungen im openCost-Kontext abstrakt abgebildet werden, um international anschlussfähig zu sein. openCost nutzt daher zur Abbildung die Einbettung in ein minimales DataCite-Schema.¹³ Gegenüber der aktuell geforderten Monitoring-Tabelle erlaubt dieser Ansatz bereits auf transparente und einfache Weise die Zuordnung zu mehr als einem Projekt und ggf. mehr als einer Fachgruppe. Beides ist über einfache Tabellen nicht darstellbar, gerade in interdisziplinären Kontexten aber geboten. Gleichzeitig demonstriert diese Implementation die Einbettung von openCost in andere Schemata beispielhaft und zeigt so die Stärke des gewählten Ansatzes.

3.4. Implementierung und Erstellung von Anwendungen

Auf Basis der OAI-Schnittstelle wurden die nötigen Ergänzungen erarbeitet und für die drei Beispielsysteme EPrints, PUB und JOIN²/Invenio implementiert.

3.4.1. Universität Regensburg

Im Publikationsserver der Universität Regensburg wurde das intern bereits vorhandene Metadatenschema zur Erfassung und Verwaltung der Publikationskosten ergänzt und an das openCost-

¹² Um die Harmonisierung der Workflows weiter zu begünstigen, haben sich DFG und FZJ entschlossen, ab 2024 ebenfalls PABA-Rechnungen als Grundlage für das DFG-Monitoring zu verwenden. Deshalb wurde die Frist für die Datenlieferung aus dem Berichtsjahr 2023 nach hinten auf den 15.8.2024 gelegt. Daher werden erst ab 2024 die für 2023 und die darauffolgenden Jahre gemeldeten Artikeldaten vollständig automatisiert von Jülich an OpenAPC übermittelt werden.

¹³ Aus dem DataCite-Schema sind konkret nur die Elemente ‚fundingReferences‘ zur Projekt- und ‚subject‘ zur Fachzuordnung notwendig, um die OAPK-Anforderungen zu erfüllen. Kostendaten werden gem. openCost dargestellt und alle Elemente in einen DataCite-Container eingebettet.

Schema angepasst. Die Daten werden, wie alle Metadaten einer Publikation, unter einer CC0-Lizenz zur freien Nachnutzung zur Verfügung gestellt. Publikationen mit assoziierten Kosten (direkt bezahlte Publikationsgebühren oder indirekte Bezahlung durch Teilnahme an einem TV, Mitgliedschaft oder Community-Geschäftsmodell) werden in der OAI-Schnittstelle¹⁴ im Metadatenformat „openCost“ und Set „openCost“ zur Verfügung gestellt.¹⁵

Unter <https://opencost.uni-regensburg.de/> steht eine frei zugängliche EPrints-Instanz zur Verfügung. Interessierte können sich einen Account erstellen und die Umsetzung des Metadatenschemas in EPrints ansehen. Die Software steht unter der freien Lizenz GPL v3 zur Nachnutzung zur Verfügung.

3.4.2. Universität Bielefeld

An der Universität Bielefeld werden die Publikationsgebühren primär über das Bibliothekssystem (ExLibris Alma) verwaltet. Ein Report liefert DOIs und Kostendaten, die mithilfe von CrossRef mit bibliografischen Metadaten angereichert und im institutionellen Repositorium PUB gespeichert werden, das auf LibreCat basiert.¹⁶ Im Rahmen des Projekts wurde die OAI-PMH-Schnittstelle von PUB so erweitert, dass die entsprechenden Kostendaten der Universität Bielefeld im openCost-Format durch OpenAPC geharvestet werden konnten.¹⁷

Datenmeldungen an OpenAPC wurden bislang i. d. R. mittels csv-Dateien über E-Mail oder GitHub durchgeführt.¹⁸ Durch ein OAI-Harvesting-Verfahren können die Daten auf automatisiertem Weg und somit stets auf dem letzten Stand des Datenlieferanten eingesammelt werden. Die Standardisierung auf das openCost-Schema legt die Grundlagen für eine breite Nachnutzung. Der bisherige prototypische Harvesting-Code in OpenAPC wurde angepasst und ergänzt.¹⁹ Für die Datenlieferung aus dem OAPK-Monitoring wurde darüber hinaus eine Verarbeitung auf Dateibasis implementiert, da Jülich zu diesem Zeitpunkt noch nicht über OAI-PMH liefern konnte.²⁰

Der Hauptspeicherort für alle in OpenAPC aggregierten Daten ist das projekteigene Repository auf GitHub, wo sowohl Roh- als auch verarbeitete Daten vorgehalten und gepflegt werden.²¹ Daneben betreibt OpenAPC einen OLAP-Server²², der als Backend für die Treemap-Visualisierung

¹⁴ Siehe <https://epub.uni-regensburg.de/cgi/oai2?verb=ListRecords&metadataPrefix=openCost&set=openCost>.

¹⁵ Während des Projektverlaufs stellte sich heraus, dass lediglich ein Set mit dem Namen „openCost“ für die Abbildung in der OAI-Schnittstelle benötigt wird, um alle Kostendaten zu liefern. Eine Filterung nach spezifischen Kostenarten kann dann beim Aggregator durchgeführt werden.

¹⁶ Siehe <https://pub.uni-bielefeld.de/>.

¹⁷ Siehe <https://github.com/OpenAPC/openapc-de/releases/tag/v4.121.10-7-6>.

¹⁸ Dies setzt eine aktive Lieferung durch die Institutionen voraus und führt zu einer Verzögerung der Bereitstellung.

¹⁹ Siehe <https://github.com/OpenAPC/openapc-de/commit/33ac814956f>.

²⁰ Hierzu wurde ein zusätzliches Skript entwickelt, das die XML-Dateien so transformiert, dass sie problemlos in OpenAPC verarbeitet werden können. Vgl. https://github.com/OpenAPC/openapc-de/blob/master/python/transform_opencost_xml.py.

²¹ Siehe <https://github.com/OpenAPC/openapc-de>.

²² Siehe <https://olap.openapc.net/>.

gen²³ dient und zugleich auch als öffentliche API für den maschinellen Zugriff auf die OpenAPC-Daten fungiert. Genutzt wird dieser sowohl zur Einbindung von OpenAPC-Daten in Repositorien (vgl. 3.4.4.) als auch den EZB-Erweiterungen (vgl. 3.5.). Der OLAP-Server wurde hierzu um ISSN und ROR-ID als jeweils abfragbares Metadatum erweitert.

Um Repositorien die Darstellung von Kostenübernahmen durch andere Einrichtungen transparent zu ermöglichen, wurde eine gezielte Abfrage von DOIs realisiert, die die für diese Publikation angefallenen Kosten und die jeweils zahlende Einrichtung liefert.²⁴

OpenAPC hat sich bislang auf die Darstellung von APC-Kosten konzentriert. Die Umstellung auf das openCost-Schema ermöglicht es, Zusatzkosten wie page oder color charges strukturiert und standardisiert zu erfassen. Im Rahmen des Projekts wurden Kostendaten von den drei JOIN²-Institutionen DESY, FZJ und GSI Helmholtzzentrum für Schwerionenforschung, die bereits Zusatzkosten erfassen, geharvestet und ausgewertet.²⁵ So gewonnene Erfahrungen zeigen, dass die Aggregation zusätzlicher Kostenarten sinnvoll und ohne größere technische Probleme möglich ist und die entsprechende Funktionalität während der Restdauer des Projekts planmäßig in den Regelbetrieb von OpenAPC übernommen werden kann.²⁶

3.4.3. DESY (JOIN²)

Die JOIN²-Repositorien können seit 2016 Publikationskosten mit den entsprechenden Publikationen nachweisen. Mit den ersten Definitionen der Monitoring-Anforderungen für OAPK wurde das verwendete Publikationskostenmodul zunächst von DESY um eine Ausgabe der geforderten Monitoringtabelle ergänzt, sodass die notwendigen Meldungen von JOIN²-Partnern weitestgehend automatisch erzeugt werden können. Hier gewonnene Erkenntnisse flossen in die Entwicklung der openCost-Formate ein. Gleichzeitig wurde von DESY für JOIN² die Darstellung der erfassten Daten im openCost-Format realisiert und der Code allen Partner*innen zur Verfügung gestellt. Weiter wurden verschiedene Automatismen in der Erfassungsoberfläche implementiert, die es den Bearbeitenden erleichtern, die Zugehörigkeit zu TV zu erkennen und korrekt anzugeben.²⁷ Parallel wurde die Zuordnung von Publikationen zu einzelnen TV generalisiert, indem Verträge als eigenständige, kostentragende Entitäten im Repositorium abgebildet werden. Diese stellen auch eine Verbindung zur ESAC-Registry her. Die interne Abbildung basiert hierbei auf dem von openCost vorgeschlagenen Format und einigen technische nötigen Ergänzungen. Dies zeigt, dass die openCost-Definition praxisgerecht ist. Ferner wurde die Erfassungsmaske um eine halb-

²³ Siehe <https://treemaps.openapc.net/>.

²⁴ Vgl. <https://github.com/OpenAPC/openapc-olap/blob/master/HOWTO.md#doi-lookup>

²⁵ Vgl. <https://github.com/OpenAPC/openapc-de/tree/opencost-prototype?tab=readme-ov-file#additional-costs>

²⁶ Die zusätzlichen Kostendaten wurden in den neu geschaffenen Datensatz „apc_additional_costs“ aufgenommen (https://github.com/OpenAPC/openapc-de/blob/opencost-prototype/data/apc_additional_costs.csv). Der Ansatz, diese Kosten in einer separaten Tabelle zu speichern und über die DOI mit dem APC-Kerndatensatz zu verknüpfen, greift nicht in die bestehenden OpenAPC-Daten ein, sodass die langfristige Datenkonsistenz nicht beeinträchtigt wird.

²⁷ Es ist bereits im Code vorgesehen, die internen Listen bei Verfügbarkeit durch die EZB-API zu ersetzen.

automatische Erkennung der DFG-Förderkriterien sowie der Kriterien zur Finanzierung aus entsprechenden Publikationsfonds erweitert.

Die von JOIN² erzeugten Daten wurden während der Implementation auf Konformität zum openCost-Schema kontinuierlich geprüft. So sind zahlreiche Beispieldateien durch automatische Datenabfragen und Validierungen gegen das Schema entstanden und jeweils aktualisiert worden. Während der gesamten Projektlaufzeit haben die JOIN²-Instanzen stets zeitnah die notwendigen Anpassungen am openCost-Format nachvollzogen.

Wie im Antrag formuliert, steht bei JOIN² eine größere Systemmigration an. Zum Zeitpunkt der Antragstellung wurde davon ausgegangen, dass diese von Invenio 1.x zu Invenio 3.x erfolgen wird. Diese anvisierte Migration hat sich zunächst deutlich verzögert, gleichzeitig hat sich Invenio 3.x nur bedingt als tauglich für die Anforderungen erwiesen. Nach intensiven Diskussionen wird das Nachfolgesystem daher auf Basis von MyCoRe MIR realisiert. Folgerichtig arbeitet DESY derzeit zusammen mit den Partner*innen von MyCoRe an der Implementation von openCost auf dieser Basis.²⁸

Die von openCost eingeleitete Automatisierung ermöglicht einigen Partner*innen nun die aktive Teilnahme an OpenAPC.²⁹ Die Automatisierung der Datenlieferung aus JUSER, der JOIN²-Instanz des FZJ, welches bisher manuell Daten an OpenAPC liefert, konnte zur Prüfung der OAI-Set-Definition und Ausgabeformate auf Vollständigkeit herangezogen werden. Ein manueller Abgleich, der via Harvesting gelieferten Daten mit den bisher in OpenAPC vorliegenden, ergab eine große Übereinstimmung. Jedoch konnten einige Lücken, die in der manuellen Übertragung aufgrund der Lieferung zu bestimmten Zeitpunkten entstanden, ergänzt und fehlerhaft gemeldete Kosten korrigiert werden. Bei der genauen Analyse der manuellen Lieferungen wurden folgende Aspekte als problematisch identifiziert:

- Abfragezeitpunkt³⁰
- Manuelle Datenkorrekturen³¹
- Korrekturen durch automatische Prozesse³²

²⁸ Die Realisation erfolgt in enger Abstimmung mit der Verbundzentrale Göttingen und weiteren bestehenden MyCoRe-Anwendern mit dem erklärten Ziel MyCoRe „openCost-Ready“ zu machen.

²⁹ Z. B. RWTH Aachen <https://openapc.github.io/general/openapc/2023/01/26/rwth/> bzw. DESY <https://openapc.github.io/general/openapc/2023/02/01/desy/>

³⁰ Der manuelle Report wird zu einem bestimmten Zeitpunkt und nach einfach definierbaren Kriterien generiert (z. B. Publikationsjahr) erstellt. Z. T. sind zum Zeitpunkt des Datenabzugs einzelne Publikationen noch in Bearbeitung oder werden erst später nachgetragen und so nicht berücksichtigt. Die Identifizierung dieser Publikationen für die Folgelieferung ist in der Praxis zeitaufwendig und fehleranfällig.

³¹ Bei internen Auswertungen werden z. T. zu späteren Zeitpunkten Fehler entdeckt und korrigiert. In der Praxis betraf dies z. B. die Erfassung von Brutto- statt Nettobeträgen. Diese fallen in den Plausibilitätskontrollen, insb. bei vorsteuerabzugsfähigen Einrichtungen, nicht auf.

³² Hier gilt das gleiche wie für manuelle Korrekturen auf Basis einzelner Datensätze. In der Praxis ist dieser Fall vergleichsweise selten, betrifft dann aber häufig eine größere Menge. So zeichnet z. B. das in Jülich verwendete Portal OABLE alle OA-Gebühren als „APC“ aus, sowohl für Gold als auch Hybrid-OA. Die betroffenen Datensätze wurden maschinell korrigiert und die Änderungen im Harvesting wie erwartet korrekt propagiert.

Diese Probleme werden durch das vorgeschlagene Harvestingverfahren eliminiert, da beim OAI-Harvesting Datensätze dann neu exponiert werden, wenn sich diese auf Systemebene ändern. Über dieses Verfahren sind auch Datenkorrekturen möglich, die einen über lange Zeit gewachsenen Korpus betreffen, wie ein Anwendungsfall aus der JOIN²-Instanz RWTH Aachen aufzeigen konnte.³³

3.4.4. Integration der OpenAPC-Informationen in Repositorien

Die Repositorien der Projektpartner*innen nutzen OpenAPC-Daten zur öffentlichen Darstellung von Kosten und zahlenden Einrichtungen auf den jeweiligen Detailseiten. So werden über OpenAPC auch Kostendaten bei Einträgen angezeigt, die von anderen Forschungseinrichtungen bezahlt wurden. Hierbei wurden prototypisch zwei Ansätze verfolgt. Die Universitätsbibliothek Regensburg nutzt einen Datenabzug von OpenAPC und ergänzt die Metadaten ihrer Publikationen. Die Daten stehen im Repository für Auswertungen daher direkt zur Verfügung. So kann ermittelt werden, in welcher Höhe und wie häufig eine Einrichtung von anderen profitiert. Die JOIN²-Institutionen und die Universitätsbibliothek Bielefeld nutzen die DOI einer Publikation zur direkten Abfrage des OLAP-Servers. Eine Aktualisierung von OpenAPC hat hier unmittelbar eine Aktualisierung der Anzeige zur Folge. Es wird keine Änderung am Internformat des Repositories notwendig, sodass dieser Ansatz breiter anwendbar ist. Auswertungen wie in Regensburg müssen hingegen über externe Werkzeuge erfolgen, die auf den OpenAPC-Datensatz zugreifen. Beiden Fällen ist gemein, dass im Sinne größtmöglicher Transparenz eine häufige und zeitnahe Aktualisierung der OpenAPC-Daten erstrebenswert ist, wie sie durch das von openCost favorisierte OAI-Harvesting ermöglicht wird.

3.5. Erweiterungen der Elektronischen Zeitschriftenbibliothek

Die EZB wird im Rahmen des Projekts umfangreich erweitert, um erwartete Publikationskosten auf der Zeitschriftenebene transparent darstellen und Informationen zu deren Übernahme anbieten zu können. Dies unterstützt Forschende bei der kostenbewussten Abwägung verschiedener Publikationsmöglichkeiten und Bibliotheksmitarbeitende, die somit Informationen ihrer Einrichtung zur Kostenübernahme einfach und zielgerichtet bereitstellen und mit geringem Aufwand pflegen können. Die Konzeption für die Erweiterung der EZB-Administration zur Eingabe von Informationen zur Übernahme von Publikationskosten ist abgeschlossen und wurde im Rahmen eines Workshops präsentiert³⁴ [10]. Zur feingranularen Darstellung der Kostenübernahmebedingungen wurden die EZB-Zeitschriften um Kategorien³⁵ erweitert, die entweder automatisiert oder manuell ver-

³³ Hier hatte die Bibliothek über das Kostenmodul für einen längeren Zeitraum Listenpreise für Publikationen ergänzt, um die nötige Ausstattung eines Publikationsfonds anhand der in JOIN² vorhandenen Monitoringtabelle abzuschätzen. Aus Sicht von OpenAPC sind diese Listenpreise als falsch anzusehen, da das Ziel die Erfassung echter, geprüfter und bezahlter Kosten ist. Sobald die fehlerhaften Daten korrigiert waren, wurden die fälschlich gemeldeten Kosten als „gelöscht“ exponiert und konnten so exakt aus dem OpenAPC-Korpus entfernt werden.

³⁴ Sippl, Weisheit (2022). Spreading Publication Cost Information with the Electronic Journals Library (EZB). DOI: 10.3204/DESY-PROC-2023-01/09, in: [9].

³⁵ Übersicht Zeitschriftenkategorien, siehe <https://ezb.ur.de/categories>.

geben werden, z. B. Open Access, Mirror Journal, Diamond Open Access Journal, Subscribe to Open.³⁶ Eine erste Nachnutzung erfolgte in Zusammenarbeit mit der DGPhil zur Auflistung von OA-Zeitschriften aus dem Fachbereich Philosophie.³⁷

Die Erfassung von Informationstexten zur Publikationskostenübernahme abhängig von Zeitschriftenkategorie und Einrichtung befindet sich in der technischen Umsetzung.³⁸ Ein Entwurf³⁹ zur Anzeige der Informationen zum Publizieren in der Benutzeroberfläche wurde in enger Abstimmung mit der Community erarbeitet.⁴⁰ Derzeit wird an der technischen Umsetzung gearbeitet. Die Veröffentlichung der erweiterten Benutzeroberfläche erfolgt im Mai 2024, zusammen mit der angepassten XML-Ausgabe zur automatisierten Nachnutzung. Die Konzeption für die Erweiterung der Administration mit einer Eingabemöglichkeit von Vereinbarungen mit Verlagen bzw. Mitgliedschaften ist fortgeschritten, die technische Umsetzung steht noch aus.

Ein weiterer zentraler Punkt ist die Einbindung von Durchschnittskosten aus OpenAPC in die Benutzeroberfläche der EZB. Hierzu wurde das Datenmodell der EZB um ISSN-L und ROR-ID⁴¹ ergänzt, was es erlaubt, Durchschnittskosten einer Zeitschrift sowohl global als auch einrichtungsspezifisch darzustellen.

Zudem wurde die EZB um Datenfelder aus DOAJ erweitert.⁴² Sofern eine Zeitschrift im DOAJ gelistet ist, wird diese automatisiert mit dieser Kategorie versehen, woraus die Nutzenden Informationen zu Zeitschriftenqualität und Förderungsmöglichkeiten ableiten können.

In der EZB können den Zeitschriftenkollektionen aus TV nun außerdem die ESAC-ID⁴³ (sofern vorhanden⁴⁴) zugeordnet werden. Zudem wurde eine eigene Ansicht für alle in der EZB hinterlegten TV geschaffen.⁴⁵ Derzeit wird eine REST-API konzipiert und implementiert, die diese Informationen institutions- und zeitschriftenspezifisch zur Nachnutzung bereitstellt (vgl. 3.4.3.).

3.6. Öffentlichkeitsarbeit, Dokumentation und Koordinierung des Projekts

Ein zentraler Aspekt der Projektarbeit liegt im Austausch mit der Fachcommunity. Diese sollte bereits in einem frühen Stadium in die Diskussion und die Erarbeitung des Schemas eingebunden

³⁶ Die Definition und Veröffentlichung weiterer Kategorien sind zukünftig vorgesehen.

³⁷ Siehe <https://www.dgphil.de/die-dgphil/open-access-zeitschriften-1>.

³⁸ Das Konzept hierzu wird auf der BiblioCon2024 erstmalig einem breiten Publikum vorgestellt.

³⁹ Siehe <https://ezb.ur.de/aw2023/EZB-Workshop%20Entwurf%20Detailseite.pdf>.

⁴⁰ Kernelement dieses Entwurfs ist die Bündelung der Informationen zum OA-Publizieren in einem GUI-Element. Der Entwurf wurde im Austausch mit der Community i. R. eines Online-Workshops am 15.11.2023 mit knapp 100 Teilnehmenden diskutiert [17].

⁴¹ Zum 21.3.2024 haben dies bereits 171 Einrichtungen getan.

⁴² Der Umfang der Anzeige wird eng mit der Community abgestimmt.

⁴³ Zum 21.3.2024 sind 36 EZB-Kollektionen als TV markiert und 48 EZB-Kollektionen mit einer ESAC-ID versehen worden.

⁴⁴ Eine (Nach-)Nutzung dieser Funktion erfolgt zurzeit z. B. für die DEAL-Zeitschriften, die in der EZB neben der Auszeichnung als Teil eines TV eigenes gekennzeichnet werden sollen, sodass wiederum entsprechende Informationen zur Kostenübernahme bereitgestellt werden können.

⁴⁵ Siehe <https://ezb.ur.de/services/transformative>.

werden, um möglichst viele Nutzungsszenarien abzudecken sowie evtl. fehlende Elemente oder Ungenauigkeiten frühzeitig zu erkennen und zu beheben. In diesem Kontext ist der Expert*innen-Workshop und der enge Austausch mit Projekten wie AT2OA2, Transform2Open und dem OAPK-Monitoring sowie das Engagement in Arbeitsgruppen des oa.networks [18, 22] zu sehen.

Die Fachcommunity wurde über die Projektwebseite⁴⁶ und die openCost-Mailingliste⁴⁷ sowie weitere gängige Mailverteiler regelmäßig über den Projektstand informiert (siehe 4.). Außerdem richtete openCost mehrere Online-Veranstaltungen aus, um Feedback zum aktuellen Stand des Metadatenschemas zu erhalten. Auf dieser Grundlage wurde das Schema anschließend überarbeitet.

Daneben wurde in verschiedenen Fachzeitschriften sowie durch Vorträge und andere Beiträge auf diversen (inter-)nationalen Tagungen, Kongressen und in Fokusgruppen über das Projekt informiert (siehe Beiträge [1 – 8, 10 – 13]). Ferner wurde auf den Open Access Tagen 2023 ein Posterbeitrag [10] zu den neuen Funktionen der EZB präsentiert, die im Rahmen des Projekts umgesetzt werden. Der Beitrag wurde mit dem Preis in der Kategorie „am meisten gelernt“ ausgezeichnet.⁴⁸ Auch auf dem renommierten Geneva Workshop on Innovations in Scholarly Communications (OAI13) wurde openCost im Rahmen eines Posterbeitrags [11] vorgestellt und im Wettbewerb mit dem zweiten Platz⁴⁹ ausgezeichnet. 2024 beteiligt sich openCost mit Beiträgen an diversen (inter-)nationalen Konferenzen, so z. B. dem Open Science Festival, der Open Repository Conference in Göteborg, der 112. BiblioCon2024, den Open Access Tagen 2024 und der PIDfest Conference in Prag sowie mit einem Thementisch am DSpace Praxistreffen.

Während des gesamten Projektzeitraums fand ein intensiver Austausch der Projektpartner*innen statt. Neben einem wöchentlichen Jour fixe via Zoom fand außerdem einmal im Monat ein großes Online-Projektgruppentreffen statt, an dem zudem Kooperationspartner*innen vom FZJ sowie andere Kooperationspartner*innen⁵⁰ in wechselnder Besetzung teilnahmen. So wurde die Möglichkeit geschaffen, mit anderen Projekten und Stakeholdern in Austausch zu treten und Anknüpfungspunkte herauszuarbeiten. Außerdem wurden pro Jahr zwei Projektgruppentreffen in Präsenz organisiert. Im geplanten Abschlussworkshop in Regensburg, der von 7. bis 9.10.2024 stattfinden wird, sollen die Ergebnisse und Nachnutzungsmöglichkeiten der gesamten Community vorgestellt werden.

⁴⁶ Projekt openCost, siehe www.opencost.de.

⁴⁷ Siehe <https://www-mailman.uni-regensburg.de/mailman/listinfo/opencost.info>.

⁴⁸ Vgl. <https://www.opencost.de/aktuelles/preis-oat23/>.

⁴⁹ Vgl. <https://oai.events/oai13/posters/> und <https://www.opencost.de/aktuelles/posterwettbewerb-oai13/>.

⁵⁰ U. a. fand ein Austausch mit Vertreter*innen von Transform2Open, AT2OA2, DataCite, OA Datenpraxis, PANTER, Markus Putnings (FAU Erlangen-Nürnberg) und Dr. Ulrich Herb (DSpace) statt.

4. Öffentlich zugängliche Projektergebnisse

Die Entwicklung der relevanten Schemata sowie deren Dokumentation wird in einem GitHub Repository⁵¹ zur Verfügung gestellt. Neben einer tabellarischen Beschreibung des Schemas enthält das Repository auch FAQs, eine XSD-Datei zur formalen Validierung und einen Ordner mit validen Beispieldateien. Über die Funktion der GitHub Issues kann sich die openCost-Community an der laufenden Diskussion beteiligen, Anregungen und Verbesserungsvorschläge abgeben sowie Fragen zum aktuellen Schema stellen, was von Seiten des Projekts ausdrücklich erwünscht ist.⁵²

4.1. Publikationen mit wissenschaftlicher Qualitätssicherung

- [1] Bartlewski, Julia et. al. (2023). The openCost project: Integrating publication cost data into institutional repositories. Open Repositories 2023, OR 2023, Stellenbosch, Südafrika, 12.–15.6.2023. DOI: 10.3204/PUBDB-2023-03959.
- [2] Bartlewski, Julia, und Broschinski, Christoph (2023). Metadatenchema für die Erfassung und Übertragung von Kostendaten: Bericht aus dem „openCost“-Projekt. O-Bib. Das Offene Bibliotheksjournal / Herausgeber VDB, 10(4), 1–13. <https://doi.org/10.5282/o-bib/5971>.
- [3] Bartlewski, Julia und Broschinski, Christoph (2022). Erweiterung der OpenAPC-Infrastruktur im Rahmen von openCost. Open Access Tage 2022, Bern, Schweiz, 19.–21.9.2022. DOI: 10.3204/PUBDB-2023-03640.
- [4] Deinzer, Gernot (2023). Transparente Publikationskosten: Ein Metadatenchema zum Austausch von Kostendaten. 1. Österreichischer Bibliothekskongress, Innsbruck, Österreich, 2.–5.5.2023. DOI: 10.3204/PUBDB-2023-06102.
- [5] Schweighofer, Bianca (2023). openCost: Kostentransparenz durch Bereitstellung von Publikationskosten. Überlegungen zum Datenaustausch und mögliche Anwendungen. 1. Österreichischer Bibliothekskongress, Innsbruck, Österreich, 2.–5.5.2023. DOI: 10.3204/PUBDB-2023-03631.
- [6] Schweighofer, Bianca (2022). Das Projekt openCost auf dem 8. Bibliothekskongress 2022. O-Bib. Das Offene Bibliotheksjournal / Herausgeber VDB, 9(4), 1–10. DOI: <https://doi.org/10.5282/o-bib/5862>.
- [7] Schweighofer, Bianca (2022). Expert*innensitzung „Total Cost of Publishing“. 8 Bibliothekskongress 2022, Leipzig, 31.5.–2.6.2022. DOI: 10.3204/PUBDB-2023-03633.
- [8] Schweighofer, Bianca (2022). Strategie zur Erfassung der „Total Cost of Publishing“. 8 Bibliothekskongress 2022, Leipzig, 31.5.–2.6.2022. DOI: 10.3204/PUBDB-2024-01543.
- [9] Schweighofer, Bianca und Wagner, Alexander, Hrsg. (2023). openCost on the road to publication cost transparency: proceedings, expert workshop, Hamburg, 5.–7.10.2022, Hamburg: Verlag Deutsches Elektronen-Synchrotron DESY. DOI: 10.3204/PUBDB-2023-02961.
- [10] Sippl, Colin (2023). openCost: Transparenz von Publikationskosten durch neue Funktionen in der Elektronischen Zeitschriftenbibliothek (EZB). Open-Access-Tage 2023, Berlin, 27.–29.9.2023. DOI: 10.5281/ZENODO.8326031.

⁵¹ GitHub Repository des Projekts openCost, siehe <https://github.com/opencost-de/opencost>.

⁵² Diese Möglichkeit wurde von der Community sehr gut angenommen (Stand Januar 2024 wurden 49 Issues mit Vorschlägen bzw. Rückfragen zum openCost-Schema erstellt, zu denen weiterhin 43 Kommentare abgegeben wurden). Um „GitHub-Anfänger*innen“ die Einstiegshürde zu nehmen, wurde zudem die Möglichkeit geschaffen, der Projektkoordinatorin Feedback per E-Mail zu senden, die diese dann in Issues umwandelte. I. R. d. der Online-Veranstaltungen wurde zudem jeweils eine kurze Einführung in den Umgang mit GitHub-Issues gegeben.

- [11] Sitek, Dagmar; Wagner, Alexander, et. al. (2022). Open-Access-Publikationskosten @ JOIN². Open Access Tage 2022, Bern, Schweiz, 19.–21.9.2022. DOI: 10.3204/PUBDB-2022-06677.
 - [12] Stein, Lisa-Marie et. al. (2023). The openCost project: Integrating publication cost data into institutional repositories. OAI13 – The Geneva Workshop on Innovations in Scholarly Communication, online/Geneva, Schweiz, 4.–8.9.2023. DOI: 10.5281/ZENODO.8307954.
 - [13] Stein, Lisa-Marie et. al. (2023). Transparent publizieren mit openCost: Neuigkeiten zur technischen Kostendateninfrastruktur. Open-Access-Tage 2023, Berlin, 27.–29.9.2023. DOI: 10.3204/PUBDB-2023-06095.
 - [14] Stein, Lisa-Marie et. al. (2023). Publication cost transparency within the Helmholtz Community and beyond: openCost metadata for different publication types – from APCs up to research data. HMC Conference: Better Research Through Better Metadata, HMC 2023, online, 10.–12.10.2023. DOI: 10.3204/PUBDB-2023-06134.
- 4.2. Weitere Publikationen und öffentlich gemachte Ergebnisse
- [15] Bartlewski, Julia (2024). Transparentes Kostenreporting mit OpenAPC. Erste Schritte zum Informationsbudget: Publikationskosten erfassen und sichtbar machen, Workshop open-access.network, online, 7.2.2024. DOI: <https://zenodo.org/records/10649830>.
 - [16] Deinzer, Gernot (2024). Erfassung von Publikationskosten an der Universität Regensburg: der (lange) Weg zu einer zentralen Rechnungsbearbeitung. Erste Schritte zum Informationsbudget: Publikationskosten erfassen und sichtbar machen. Workshop open-access.network, online, 7.2.2024. DOI: <https://doi.org/10.5281/zenodo.10650327>.
 - [17] Schirmer, Vicky und Weisheit, Silke (2024). Auswertung des EZB-Workshops zur Überarbeitung der Zeitschriftendetailseite am 15.11.2023. DOI: <https://doi.org/10.5283/epub.55525>.
 - [18] Schweighofer, Bianca (2023). Das Projekt openCost. Kostentransparenz durch offene Bereitstellung von Publikationskosten. Fokusgruppe „Informationsbudget“, online, 1.6.2023. DOI: 10.3204/PUBDB-2023-03910.
 - [19] Schweighofer, Bianca und Wagner, Alexander (2022). Das Projekt openCost. Jubiläumsveranstaltung „25 Jahre EZB & 20 Jahre DBIS“, Regensburg, 13.–14.9.2022. DOI: 10.3204/PUBDB-2022-04988.
 - [20] Stein, Lisa-Marie (2024). openCost & DESY: zentrale Erfassung und Nachnutzung von Kostendaten im Repositorium. Erste Schritte zum Informationsbudget: Publikationskosten erfassen und sichtbar machen. Workshop open-access.network, online, 7.2.2024. DOI: <https://zenodo.org/records/10650661>.
 - [21] Stein, Lisa-Marie (2023). openCost: Automatisierte, standardisierte Lieferung und offene Bereitstellung von Publikationskosten und Verlagsvereinbarungen. Monitoring, Informationsbudget & Alma. Workshop der Landesinitiative openaccess.nrw, Essen, 20.–23.4.2023. DOI: 10.3204/PUBDB-2023-03641.
 - [22] Stein, Lisa-Marie (2023). openCost in der Praxis: Kostentransparenz in wissenschaftlichen Infrastrukturen am Beispiel von DESY. Fokusgruppe „Kostenkontrolle“ – open access network, online, 30.11.2023. DOI: 10.3204/PUBDB-2023-07538.
 - [23] Stein, Lisa-Marie (2022). openCost: Automatisierte, standardisierte Lieferung und offene Bereitstellung von Publikationskosten und Verlagsvereinbarungen. Helmholtz Open Science Forum, Berlin, 14.9.2022. DOI: 10.3204/PUBDB-2023-03634.