

# Metadatenchema für die Erfassung und Übertragung von Kostendaten

## Bericht aus dem „openCost“-Projekt

### Ausgangslage

Das Projekt openCost ist im Kontext der Open-Access-Transformation angesiedelt.<sup>1</sup> Unter Open-Access-Transformation wird der aktuelle Wandel des akademischen Publikationsbetriebs von Subskriptionsmodellen hin zu einem freien Zugriff auf Informationen verstanden. Publizieren im Open Access ist dabei als politisches Ziel klar formuliert, indem auch Förderorganisationen ihre Unterstützung daran knüpfen.<sup>2</sup>

Um Fördermittel zu erhalten, wird das Monitoring der eigenen Publikationskosten immer mehr zur Voraussetzung, wie beispielsweise im Zuge des Förderprogramms „Open Access Publikationskosten“ durch die Deutsche Forschungsgemeinschaft (DFG). Mit dem Programm fördert die DFG wissenschaftliche Einrichtungen mit finanziellen Zuschüssen für Kosten, die bei der Veröffentlichung von Open-Access-Publikationen entstehen. Als übergeordnetes Ziel formuliert die DFG für ihr Programm, „eine Strukturanpassung der Finanzierungströme für die Open-Access-Transformation zu ermöglichen und dabei die Transparenz hinsichtlich der Kosten für die Open-Access-Publikation von wissenschaftlichen Ergebnissen zu verbessern.“<sup>3</sup>

Dies geht konform mit den aktuellen Debatten zum Thema „Informationsbudget“. In seinen „Empfehlungen zur Transformation des wissenschaftlichen Publizierens zu Open Access“ vom Januar 2022 regt der Wissenschaftsrat die Einführung eines Informationsbudgets an wissenschaftlichen Einrichtungen an.<sup>4</sup> Das Informationsbudget dient dabei als Wissens- und Steuerungsinstrument, das die Open-Access-Transformation unterstützen soll. Ziel ist es, Mittel und Kosten ganzheitlich zu betrachten und idealerweise zentral zusammenzuführen. Heinz Pampel definiert das Informationsbudget als „alle Einnahmen und Ausgaben für die wissenschaftlichen Informationen“.<sup>5</sup> In seinem Bericht „Das Informationsbudget – Konzept und Werkstattbericht“ aus dem Jahr 2022 präsentiert Bernhard Mittermaier ein Schalenmodell des Informationsbudgets, in dem er 26 Elemente vorschlägt, die als Komponenten des Informationsbudgets verstanden werden sollen. Dazu zählen u.a. Ausgaben für

1 openCost - Automatisierte, standardisierte Lieferung und offene Bereitstellung von Publikationskosten und Verlagsvereinbarungen, <<https://www.opencost.de/>>, Stand: 10.07.2023. – Dieser Beitrag beruht auf dem Vortrag „Metadatenchema für die Erfassung und Übertragung von Kostendaten – Bericht aus dem „openCost“-Projekt“ am 25.05.2023 anlässlich der 111. BiblioCon in Hannover, Abstract: <<https://dbt2023.abstractserver.com/program/#/details/presentations/131>>, Stand: 10.07.2023.

2 open-access.network. Open-Access-Transformation, <<https://open-access.network/informieren/politische-rahmenbedingungen/open-access-transformation>>, Stand: 10.07.2023.

3 DFG-Förderprogramm: „Open-Access-Publikationskosten“, <[https://www.dfg.de/foerderung/programme/infrastruktur/lis/lis\\_foerderangebote/open\\_access\\_publikationskosten/](https://www.dfg.de/foerderung/programme/infrastruktur/lis/lis_foerderangebote/open_access_publikationskosten/)>, Stand: 10.07.2023.

4 Wissenschaftsrat: Empfehlungen zur Transformation des wissenschaftlichen Publizierens zu Open Access, 2022. Online: <<https://doi.org/10.57674/tyrc-vb61>>, Stand: 10.07.2023.

5 Pampel, Heinz: Auf dem Weg zum Informationsbudget. Zur Notwendigkeit von Monitoringverfahren für wissenschaftliche Publikationen und deren Kosten. Arbeitspapier, 2019, S.11. Online: <<https://doi.org/10.2312/os.helmholtz.006>>, Stand: 10.07.2023.

das OA-Publizieren, wie Article Processing Charges (APCs), aber auch Zusatzkosten wie Page- oder Colour Charges, sowie Ausgaben für Transformationsverträge.<sup>6</sup>

Die Überwachung und Steuerung der Open-Access-Transformation erfordert solide und transparente Daten zu Zahlungen, um stetig steigenden Kosten vorbeugen zu können. Zusammenfassend lässt sich festhalten, dass die Kostentransparenz im transformativen Prozess hin zu Open Access immer wichtiger wird. Doch fehlt es vielerorts noch an automatisierten und standardisierten Workflows zur Verwaltung und Meldung von Publikationskosten. Diese werden aber benötigt, um Kostentransparenz zu schaffen. An dieser Stelle setzt das Projekt openCost an.

## Kurzdarstellung des Projekts

Das Projekt openCost wird von der Deutschen Forschungsgemeinschaft (DFG) über eine Laufzeit von drei Jahren mit Beginn im Dezember 2021 gefördert (DFG-Projektnummer 457354095). Es wird von drei Projektpartner\*innen getragen. Dabei handelt es sich um die Universitätsbibliotheken Regensburg und Bielefeld sowie um die Zentralbibliothek des Deutschen Elektronensynchrotrons (DESY) in Hamburg.

Der gewählte Titelzusatz „Automatisierte, standardisierte Lieferung und offene Bereitstellung von Publikationskosten und Verlagsvereinbarungen“ umschreibt bereits in konzentrierter Form den inhaltlichen Schwerpunkt des Projekts.

## Projektziele und Bausteine

Ganz übergeordnet folgt openCost also dem Leitgedanken der Kostentransparenz mit dem Ziel, dass möglichst alle Kosten für Publikationen wissenschaftlicher Einrichtungen nachgewiesen und dokumentiert werden.

Zu diesem Zweck soll eine technische Infrastruktur entwickelt und etabliert werden, mit der Publikationskosten umfassend erfasst und über standardisierte Formate und Schnittstellen frei zugänglich abgerufen und ausgetauscht werden können.

Die konkrete Realisierung in der Praxis erfolgt über drei Bausteine. Um Kostentransparenz zu ermöglichen, ist die erste Voraussetzung, dass die Kosten in einem maschinenlesbaren und standardisierten Format erfasst werden können.

Aus diesem Grund entwickeln die Projektpartner\*innen unter Einbeziehung (inter-)nationaler Expertise ein standardisiertes Metadatenschema, mit dem Publikationskosten einer wissenschaftlichen Einrichtung in strukturierter Form erfasst, abgefragt und abgebildet werden können. Vor dem Hintergrund der Diskussion um das Informationsbudget soll das Schema dabei möglichst alle Kostenarten

---

6 Mittermaier, Bernhard: Das Informationsbudget. Konzept und Werkstattbericht, in: o-bib. Das offene Bibliotheksjournal 9 (4), 2022, S. 6ff. Online: <<https://doi.org/10.5282/o-bib/5864>>.

abdecken können. Neben Open-Access-Publikationsgebühren wie APCs beinhaltet dies auch Kosten aus Transformationsverträgen und Mitgliedschaften sowie Zusatzkosten, z.B. Page- oder Colour Charges.

Das Metadatenchema ist Grundlage für den zweiten Projektschwerpunkt, den automatisierten Austausch und die Übertragung der Kostendaten. Für diesen Harvestingprozess schlägt das openCost-Projektteam die etablierte OAI-PMH-Schnittstelle vor.<sup>7</sup> Über eine solche Schnittstelle können die Daten innerhalb einer Institution, aber auch institutionsübergreifend zusammengeführt und für verschiedene Kostenauswertungen abgerufen werden. Über OAI-PMH wird es auch Aggregatoren möglich sein, den Publikationsoutput direkt bei den Institutionen einzusammeln. Im Projekt wird dies exemplarisch für den Service OpenAPC realisiert.<sup>8</sup> OpenAPC, betrieben von der Universitätsbibliothek Bielefeld seit 2014, sammelt und veröffentlicht Datensätze über Gebühren für Open-Access-Veröffentlichungen unter einer offenen Datenbanklizenz. Neben APCs werden bislang Daten zu Transformationsverträgen sowie Kosteninformationen zu Open-Access-Monografien, sogenannte Book Processing Charges (BPCs), aggregiert. Im Rahmen von openCost wird OpenAPC erweitert, um zukünftig auch Zusatzkosten abbilden zu können.

Der dritte Schwerpunkt betrifft die Erweiterung der Elektronischen Zeitschriftenbibliothek (EZB) um Informationen zu Open-Access-Publikationskosten.<sup>9</sup> Über eine Schnittstelle wird die EZB mit OpenAPC vernetzt, sodass Kosteninformationen auf der Detailseite der jeweiligen Zeitschrift angezeigt werden können. Des Weiteren werden die Administrationsfunktionen für teilnehmende Einrichtungen erweitert. Diese sollen die Möglichkeit bekommen, Informationen zur Übernahme von Publikationskosten für bestimmte Zeitschriften in der EZB einzutragen sowie einrichtungsspezifische Vereinbarungen, wie Mitgliedschaften oder Regelungen aus Transformationsverträgen, zu erfassen und anzeigen zu lassen. Diese Informationen sollen dann über eine REST-API zur Verfügung gestellt und somit nachnutzbar angeboten werden. Die EZB könnte damit als eine zentrale Plattform hinsichtlich dieser Open-Access-Informationen für alle beteiligten Institutionen dienen. Die geplanten Erweiterungen dienen insbesondere Forschenden, die sich über Artikelkosten und mögliche Kostenübernahmen ihrer Einrichtung vorab informieren möchten.

Abbildung 1 zeigt in der Übersicht die angestrebte Vernetzung der einzelnen Dienste und visualisiert die Dynamik des Gesamtprozesses von openCost. Zunächst sind die Einrichtungen abgebildet, die openCost-Standards nutzen. Diese haben das openCost-Metadatenchema etabliert und verfügen über die technische Schnittstelle zum Austausch der Kostendaten. Im Projekt soll dies anfänglich beispielhaft für die Systeme der Universitäten Bielefeld und Regensburg sowie für DESY und die weiteren Partnerinstitutionen von „JOIN<sup>2</sup>“ geschehen. DESY ist Teil des JOIN<sup>2</sup>-Konsortiums, das zurzeit aus acht Institutionen besteht.<sup>10</sup> JOIN<sup>2</sup> bietet mehrere Instanzen aus technisch identischen Systemen. So können alle JOIN<sup>2</sup>-Institutionen von der vorgeschlagenen Schnittstelle profitieren.

7 Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), <<https://www.openarchives.org/pmh/>>, Stand: 10.07.2023.

8 OpenAPC, <<https://openapc.net/>>, Stand: 10.07.2023.

9 Elektronische Zeitschriftenbibliothek (EZB), <<https://ezb.uni-regensburg.de/>>, Stand: 10.07.2023.

10 JOIN<sup>2</sup>, <<https://join2.de/>>, Stand: 10.07.2023.

Das Projekt kann dabei auf den Erfahrungen der Projektpartner\*innen aufbauen. Die drei Institutionen haben schon erste Arbeitsabläufe für die Integration von Publikationskosten in ihren institutionellen Repositorien eingerichtet und machen diese Kostendaten bereits über OAI-PMH öffentlich zugänglich. Auf diese Weise werden die tatsächlich gezahlten Publikationskosten einzelner Einrichtungen transparent dargestellt. Für das Vorantreiben der Open-Access-Transformation ist eine Übersicht der tatsächlich geleisteten Zahlungen im Vergleich zu den Listenpreisen auf den Verlagswebseiten notwendig. Die Universität Regensburg betreibt ihren Publikationsserver, basierend auf EPrints 3, und speichert dort Rechnungen sowie Rechnungsdetails wie die bezahlten Gebühren.<sup>11</sup> Die Universität Bielefeld hat mit „PUB – Publikationen an der Universität Bielefeld“ ein institutionelles Repositorium, das auf LibreCat basiert, und aktuell Informationen zu APCs bereitstellt.<sup>12</sup> Der JOIN<sup>2</sup>-Verbund beruht auf einer gemeinschaftlichen Repositorienlösung, basierend auf Invenio. Auch hier gibt es Funktionalitäten zur Verwaltung von Publikationskosten und einige Einrichtungen, darunter das DESY, erfassen neben APCs bereits Zusatzkosten.

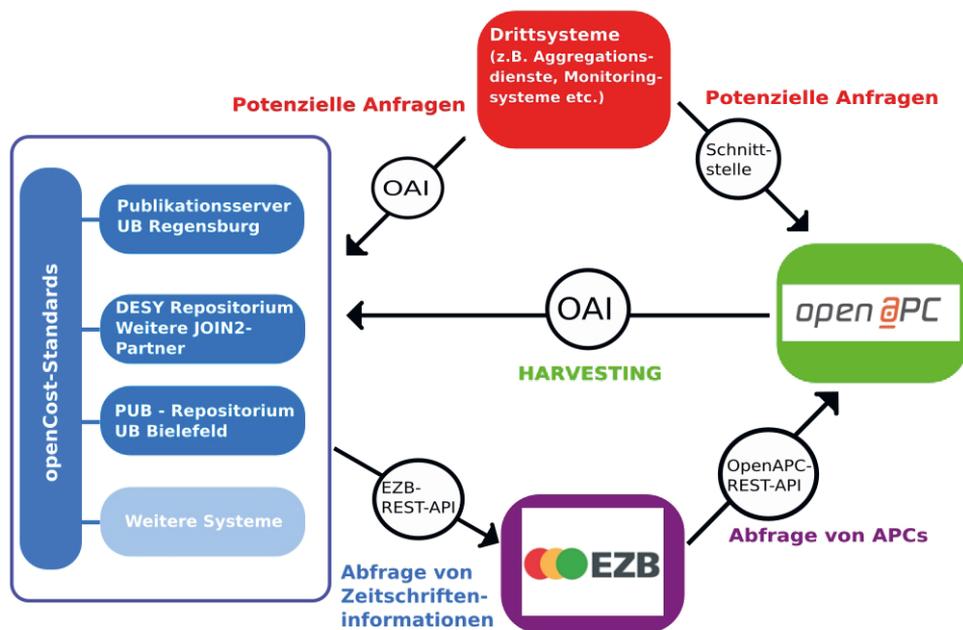


Abbildung 1: Darstellung der Vernetzung der einzelnen Dienste

Der Service OpenAPC, der künftig um die openCost-Daten erweitert wird, kann die bereitgestellten Daten direkt bei den jeweiligen Einrichtungen via OAI-PMH harvesten. Die EZB kann diese Daten dann wiederum bei OpenAPC über eine Schnittstelle abfragen. Die geplanten zusätzlichen Anzeigen zu Open-Access-Publikationskosten können dann entsprechend von den interessierten Einrichtungen

11 Publikationsserver der Universität Regensburg, <<https://epub.uni-regensburg.de/>>, Stand: 10.07.2023.

12 PUB – Publikationen an der Universität Bielefeld, <<https://pub.uni-bielefeld.de/>>, Stand: 10.07.2023.

über eine API ausgelesen werden. Via Schnittstelle könnten aber auch weitere Drittsysteme direkt bei den Institutionen Anfragen stellen und deren Publikationsausgaben harvesten oder ebenfalls die Daten bei OpenAPC für einen Kostenvergleich o.ä. abfragen. Ein solcher Datenaustausch könnte es ermöglichen, Beziehungen zu anderen Institutionen herzustellen und ein Monitoring gezahlter Gebühren zu realisieren. Sind die in den Repositorien hinterlegten Daten zudem mit gut gepflegten Metadaten angereichert, wäre eine gekoppelte Auswertung der bezahlten Gebühren nach Fachzugehörigkeit, Projektbeteiligung, Lizenzarten oder Art der Publikationsgebühr möglich. Derartige Auswertungen werden sicherlich für zukünftige Verteilungsmodelle von Geldern aus Bibliotheks-etats, Publikationsfonds oder Drittmittelzuweisungen von großer Bedeutung sein und damit auch für Förderer von großer Relevanz.

## openCost Metadatenchema

Die Entwicklung eines Metadatenchemas zum Austausch von Publikationskosten ist ein zentrales Arbeitspaket des Projekts und eine wichtige Grundlage für weitere Projektziele. Um Transparenz und einen umfassenden Überblick über gezahlte Kosten im Zusammenhang mit dem wissenschaftlichen Publizieren zu ermöglichen, genügt es nicht, nur Open-Access-Gebühren zu erfassen. Das Projekt nimmt die gesamten Kosten in den Blick, die für Publikationen aufgewendet werden (Total Cost of Publishing). Außerdem verfolgt openCost mit seinem Schema einen generischen Ansatz, sodass dieses von vielen Einrichtungen nachgenutzt werden kann und nicht an konkrete Dienste und Anwendungen gebunden ist.

## Einbindung der Community

Ein zentraler Aspekt der Projektarbeit liegt im Austausch mit der Fachcommunity. Für den Erfolg des Projekts ist es entscheidend, dass das Metadatenchema praxisbezogen ist und möglichst viele Nutzungsszenarien abdecken kann. Aus diesem Grund hat sich das openCost-Team dazu entschieden, die Community in alle Entwicklungsschritte explizit mit einzubeziehen. Deshalb wurden bereits mehrere Workshops und Veranstaltungen abgehalten, die der Erarbeitung eines standardisierten Schemas dienen.

In einem frühen Stadium des Projekts war openCost mit zwei Veranstaltungen auf dem 8. Bibliothekskongress in Leipzig im Mai 2022 vertreten, wo das Team das Projekt erstmalig einem größeren Publikum vorstellen konnte. In dem öffentlichen Hands-on Lab „Strategie zur Erfassung der ‚Total Cost of Publishing‘“ lag der Fokus auf einem Informationsaustausch mit den Teilnehmenden, um Bedarfe sowie Probleme hinsichtlich der Kostenerfassung herausarbeiten. Daneben wurden in einer internen Arbeitssitzung der neu gegründeten Expert\*innengruppe „Total Cost of Publishing“ die Ergebnisse aus diesem Workshop diskutiert und die nächsten Schritte ausgearbeitet.<sup>13</sup>

---

<sup>13</sup> Zusammenfassender Bericht: Schweighofer, Bianca: Projekt openCost auf dem 8. Bibliothekskongress 2022, in: o-bib. Das offene Bibliotheksjournal 9 (4), 2022, S. 1–10. Online: <<https://doi.org/10.5282/o-bib/5862>>.

Die gewonnen Erkenntnisse bildeten die Diskussionsgrundlage für den Workshop „openCost: the road to publication cost transparency“, der im DESY in Hamburg vom 5. bis 7.10.2022 stattfand. Dieser Meilenstein des Projekts diente zum einen der Präsentation erster Ergebnisse und zum anderen dem Wissens- und Informationsaustausch mit nationalen und internationalen Expert\*innen aus dem Bereich ‘Erfassung von Publikationskosten’. So haben Speaker\*innen aus acht verschiedenen Ländern (Deutschland, Österreich, Niederlande, England, USA, Kanada, Norwegen und Finnland) von ihren Erfahrungen berichtet. Darüber hinaus hat das openCost-Team mit den Teilnehmenden in Hands-on Labs gemeinsam an der Ausgestaltung des Metadatenschemas gearbeitet.<sup>14</sup>

Die Ergebnisse aus diesem Workshop waren dann der Ausgangspunkt für die Weiterentwicklung eines Entwurfs für ein Metadatenschema, der im Februar 2023 veröffentlicht wurde. Begleitet wurde dieser von einer ersten Kommentierungsphase und öffentlichen Online-Diskussionen. Das Feedback diente zur Feinabstimmung des Entwurfs. Dabei kristallisierten sich Handlungsfelder und Anforderungen heraus, die konkret in die Arbeit an dem Schema eingeflossen sind. Was muss diskutiert werden, wenn man solche Standards festhalten möchte? Ein Beispiel dafür sind Terminologien, mit denen man sich auseinandersetzen muss. Hier wurden aufgrund von Anregungen und Feedback aus der Community einige Elemente im Schema umbenannt. Aber auch die Diskussion um Identifikatoren hat großen Raum eingenommen. Welche primären Identifier machen Sinn und inwieweit sind bibliographische Informationen abzubilden in dem Schema? Daneben haben sich auch einige zentrale Probleme herauskristallisiert, die immer wieder benannt wurden. Wie geht man mit Sammelrechnungen, Rabatten, externen Kofinanzierungen, aber auch mit Kosten ohne konkreten Publikationsbezug, bspw. Mitgliedschaften, um?<sup>15</sup>

## Aktuelle Version des openCost-Schemas

In dem aktuellen Schema geht es zunächst um die Abbildung der Kosten für gebührenpflichtige Zeitschriftenartikel. Ein üblicher Anwendungsfall wären kostenpflichtige Artikel in Gold-Open-Access-Zeitschriften, wobei in dem Schema auch bereits Kostenelemente abgebildet werden können, die über Open-Access-Gebühren hinausgehen.

Da der Datenaustausch über OAI-PMH erfolgen soll, wird das Schema zunächst im XML-Format abgebildet. Angedacht ist, das Schema zukünftig auch im JSON-Format anzubieten. Das openCost-Team nutzt für alle Arbeiten an dem Schema ein Repository auf GitHub, da darüber Anpassungen und Erweiterungen öffentlich nachvollzogen werden können.<sup>16</sup> Das openCost-Repository enthält neben einer tabellarischen Beschreibung des Schemas auch FAQs, eine XSD-Datei zur formalen Verifikation und einen Ordner mit Beispieldateien, die die Anwendung des Metadatenschemas illustrieren.

14 Die Ergebnisse des Expert\*innenworkshops wurden als Conference Proceedings veröffentlicht: Schweighofer, Bianca; Wagner, Alexander (Hg.): Proceedings of the Expert Workshop openCost on the Road to Publication Cost Transparency, 2023. Online: <<http://dx.doi.org/10.3204/PUBDB-2023-02961>>.

15 Schweighofer, Bianca: Ergebnisse der Online-Diskussionen, in: openCost, 24.04.2023. Online: <<https://www.open-cost.de/aktuelles/ergebnisse-online-diskussionen/>>, Stand: 10.07.2023.

16 openCost Repository: GitHub - opencost-de/opencost: openCost data formats, documentation and examples, <<https://github.com/opencost-de/opencost>>, Stand: 10.07.2023.

Abbildung 2 zeigt eine Übersicht über die Struktur des aktuellen openCost-Schemas (Stand: 10. Juli 2023).



Abbildung 2: Struktur des aktuellen openCost-Schemas, Stand: 10. Juli 2023

Das Top-Element „publication“ entspricht dabei einer einzelnen Publikation. Diese sollte mithilfe eines persistenten Identifikators eindeutig identifizierbar sein. Insbesondere für Zeitschriftenartikel werden DOIs international zur Kennzeichnung verwendet. Sie haben zudem den Vorteil, dass Metadaten einfach importiert werden können, weshalb das Schema als primären Identifier eine DOI vorsieht. Sollte keine DOI verfügbar sein, ist im Schema ein bibliographischer Block als Alternative anzugeben.<sup>17</sup> Das Feld „primary\_identifier“ ist verpflichtend.

17 Auf die Angabe bibliographischer Metadaten sollte möglichst verzichtet und der bibliografische Block sollte nur in Ausnahmefällen genutzt werden. Falls erforderlich, sollten bibliografische Informationen auf andere Weise hinzugefügt werden. Optionen, wie die Einbettung des openCost-Schema in weitere Metadaten schemata oder die Abfrage bibliografischer Metadaten mittels OAI-ID, stehen momentan zur Diskussion.

Zusätzlich wurde die Möglichkeit geschaffen, weitere Identifier, wie eine lokale ID oder eine OAI-ID, über das Feld „secondary\_identifiers“ hinzuzufügen. Dieses ist optional, kann jedoch mehrfach vorkommen.

Über „institution“ wird die Einrichtung definiert, die für den Artikel bezahlt hat. Zur Identifizierung der Einrichtung gib es in dem Schema unterschiedliche Möglichkeiten. Idealerweise sollte sie durch einen persistenten Identifier identifiziert werden. Das Projekt-Team empfiehlt die Nutzung einer ROR-ID.<sup>18</sup> Alternativ ist es auch möglich, einen menschenlesbaren Namen einzugeben.

Für die eigentlichen Kostendaten ist das Invoice-Element vorgesehen. In diesem Datenblock können alle relevanten Informationen zu den Kosten eines Artikels abgebildet werden. Dabei korrespondiert das Element mit einer realen Rechnung. Innerhalb des Invoice-Blocks können Daten erfasst werden, die man in der Regel auf einer Rechnung finden würde. Die Invoice-Elemente dürfen mehrfach vorkommen, da es zu einer Publikation mehrere Rechnungen geben kann. Es ist verpflichtend, dass der Block mindestens einmal vorkommt. Das Feld „amount\_invoice“ gibt den monetären Betrag an, der als Gesamtsumme auf der Rechnung steht. Als zusätzliches Attribut muss die Währung angegeben werden. Über das Feld „invoice\_number“ kann optional die Rechnungsnummer angegeben werden. Das Feld „amounts\_paid“ gibt die Beträge als Zahlenwert an, die auch tatsächlich gezahlt wurden. Dieses Feld darf mehrfach vorkommen, mindestens einmal muss dieses Feld jedoch befüllt werden. Das Element „amount\_paid“ hat zwei verpflichtende Attribute, zum einen die Währung und zum anderen den Typ. Der Typ beschreibt die Kostenart. Man hat die Möglichkeit aus einem vorgeschriebenen Vokabular auszuwählen.<sup>19</sup> Des Weiteren gibt es die Möglichkeit, das Bezahldatum und/oder das Rechnungsdatum anzugeben, wobei die Nennung eines der beiden Daten verpflichtend ist. Über das Feld „creditor“ wird der Zahlungsempfänger als String (Zeichenkette) angegeben, dieses ist nicht normiert.<sup>20</sup>

Die Beispieldatei (Abb.3) zeigt, wie ein Datensatz, welcher das openCost-Schema erfüllt, aussehen könnte. Aus den Informationen im Kostenblock wird ersichtlich, dass die Rechnung in US-Dollar gestellt wurde und dass die „amount\_paid“-Elemente einzelne Posten sind, die in Euro bezahlt wurden. Darüber hinaus sind auch verschiedene Kostenarten angegeben. In diesem Fall wurden APCs von über 1.500€ bezahlt und zusätzlich wurde ein Page Charge von über 500€ berechnet.

---

18 Research Organization Registry (ROR), <<https://ror.org/>>, Stand: 10.07.2023.

19 Festes Vokabular zur Benennung der Kostentypen: gold-oa, hybrid-oa, vat, colour charge, cover charge, page charge, permission, publication charge, reprint, submission fee, payment fee oder other.

20 Für die detaillierte Dokumentation vgl:<<https://github.com/opencost-de/opencost/tree/main/doc#readme>>, Stand: 10.07.2023.

```

<opencost:publication xmlns:opencost="https://opencost.de">
  <opencost:institution>
    <opencost:institution_name type="short">desy</opencost:institution_name>
    <opencost:institution_id type="ror">https://ror.org/01js2sh04</opencost:institution_id>
  </opencost:institution>
  <opencost:primary_identifier>
    <opencost:doi>10.1364/OME.460445</opencost:doi>
  </opencost:primary_identifier>
  <opencost:secondary_identifiers>
    <opencost:id type="local">PUBDB-2022-02875</opencost:id>
    <opencost:id type="oai">oai:bib-pubdb1.desy.de:479017</opencost:id>
  </opencost:secondary_identifiers>
  <opencost:type>journal article</opencost:type>
  <opencost:cost_data>
    <opencost:invoice>
      <opencost:amount_invoice currency="USD">2173.0</opencost:amount_invoice>
      <opencost:amounts_paid>
        <opencost:amount_paid currency="EUR" type="gold-oa">1501.58</opencost:amount_paid>
        <opencost:amount_paid currency="EUR" type="page charge">546.68</opencost:amount_paid>
      </opencost:amounts_paid>
      <opencost:creditor>OSA</opencost:creditor>
      <opencost:dates>
        <opencost:date_paid>2022-06-09</opencost:date_paid>
      </opencost:dates>
    </opencost:invoice>
  </opencost:cost_data>
</opencost:publication>

```

Identifizierer für zahlende Einrichtung

Primärer (DOI) und sekundäre Identifier für die Publikation

Invoice-Block (mehrfach möglich)

Abbildung 3: Beispiel-Record im openCost Format

## Anwendungsfälle

### Anwendungsfall 1: Datenübermittlung des DFG-Monitorings an OpenAPC

Im Rahmen des DFG-Förderprogramms „Open Access Publikationskosten“ haben sich die partizipierenden Einrichtungen dazu verpflichtend, Daten zu den geförderten Publikationen öffentlich zugänglich zu machen. Mit dem Monitoring wurde die Zentralbibliothek des Forschungszentrums Jülich beauftragt.<sup>21</sup> Über das zu nutzende Jülicher Metadatenchema werden neben Publikationsmetadaten auch Informationen zu Kostendaten erfasst. Den teilnehmenden Einrichtungen wird im Ablieferungsprozess die Möglichkeit gegeben, die Daten für eine Nachnutzung durch OpenAPC freizugeben. Da es gewisse strukturelle Ähnlichkeiten zwischen dem Jülicher Schema und dem openCost-Schema gibt, soll das openCost-Schema zur Übermittlung der Daten aus dem DFG-Monitoring an OpenAPC verwendet werden. Ein erster Datentransfer vom Forschungszentrum Jülich an OpenAPC konnte bereits erfolgreich realisiert werden, wobei APC-Daten im openCost-Format übermittelt wurden.<sup>22</sup>

21 Monitoring des DFG-Förderprogramms „Open-Access-Publikationskosten“, <<https://www.fz-juelich.de/de/zb/open-science/open-access/monitoring-dfg-oa-publikationskosten>>, Stand: 10.07.2023.

22 Zum Zeitpunkt des Redaktionsschlusses war die Verarbeitung der Daten in OpenAPC noch nicht abgeschlossen. Neuigkeiten und aktuelle Entwicklungen sind über die Projekt-Webseite einsehbar: <<https://www.opencost.de/>>.

## Anwendungsfall 2: Direktes Harvesting von zusätzlichen Kostendaten

Der zweite Anwendungsfall, der sich zurzeit noch in einer Betaversion befindet, betrifft das direkte Harvesting von zusätzlichen Kostendaten durch OpenAPC. Bereits jetzt stellen einige Einrichtungen, die sich bei OpenAPC einbringen, ihre Kostendaten zu den Publikationen direkt über ihre institutionellen Repositorien zur Verfügung. Bisher handelt es sich dabei um spezialisierte Einzellösungen. OpenAPC wird daher zukünftig das openCost-Schema verwenden, um einerseits auf ein einheitliches und strukturiertes Schema zu wechseln und andererseits die Möglichkeit zu bieten, nicht nur APCs, sondern auch Zusatzkosten zu berichten. Dies wurde prototypisch umgesetzt für folgende drei Einrichtungen aus dem JOIN<sup>2</sup>-Verbund: Deutsches Elektronensynchrotron DESY, Forschungszentrum Jülich und GSI Helmholtzzentrum für Schwerionenforschung.

Diese drei Einrichtungen tracken bereits Zusatzkosten und spielen sie gemeinsam mit den Publikationsdaten direkt in ihre Repositorien ein. Über die OAI-Schnittstellen wurden diese Daten im openCost-XML-Schema bereitgestellt und von OpenAPC geharvestet. Anschließend wurden die Kostendaten auf Seiten von OpenAPC aufgesplittet (Abb. 4). Die APCs wurden dabei in den OpenAPC-Kerndatensatz übernommen<sup>23</sup> und alle übrigen Kostendaten, die OpenAPC bislang nicht aggregiert, wurden in einen neu geschaffenen Datensatz („openCost-Datensatz“) überspielt. Diese beiden Datensätze können dann über die DOI als Primärschlüssel miteinander verknüpft werden.

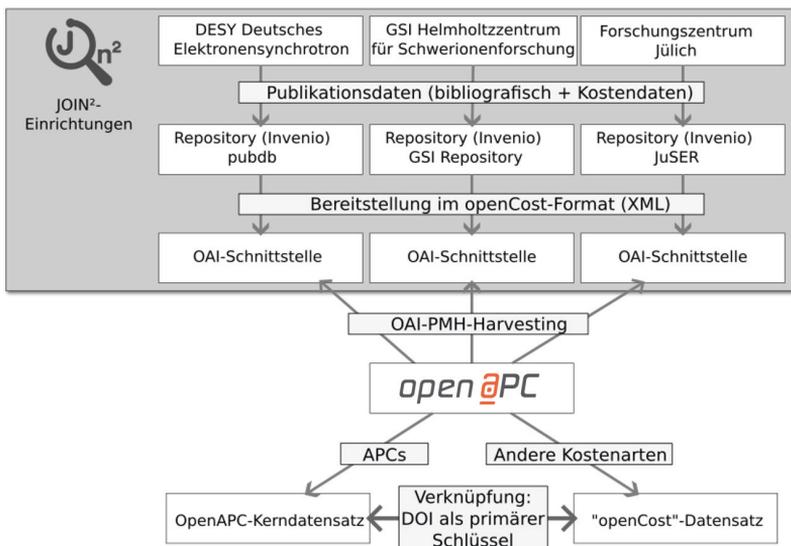


Abbildung 4: Übersicht über den OpenAPC Harvesting-Prozess von zusätzlichen Kostendaten

23 OpenAPC-Kerndatensatz, [https://github.com/OpenAPC/openapc-de/blob/master/data/apc\\_de.csv](https://github.com/OpenAPC/openapc-de/blob/master/data/apc_de.csv), Stand: 10.07.2023.

Insgesamt konnten 2320 Gold-OA-Zeitschriftenartikel geharvestet werden. Alle Artikel haben eine APC, aber es konnten auch viele zusätzliche Kostendaten identifiziert werden, wie Color Charges, Page Charges oder Reprint Fees (vgl. Tabelle 1). Eine grafische Auswertung des Datensatzes zeigt auf, wie sich diese Kostenarten über einzelne Verlage verteilen. Abbildung 5 stellt die APCs den zusammengenommenen Zusatzkosten gegenüber. Man sieht, dass APCs natürlich den größten Anteil an der Gesamtsumme ausmachen. Abbildung 6 schlüsselt die Zusatzkosten detailliert nach Kostentypen auf. Die Grafiken zeigen allerdings auch, dass es einige Fälle gibt, wie z.B. Wiley-Blackwell oder Oxford University Press, wo diese Zusatzkosten einen durchaus relevanten Anteil der Gesamtkosten ausmachen.

Tabelle 1: Übersicht über geharvestete Kostenarten in Euro

Cost type	Occurrence	Median	Sum
APC	2320	1632	4350527
Cover	9	1295	12704
Colour Charges	40	1643	72757
Page Charges	63	831	63271
Other	114	28	5900
Permission	9	189	1660
Publication Charges	15	1143	18958
Reprint	7	399	10450
Submission Fee	10	68	768

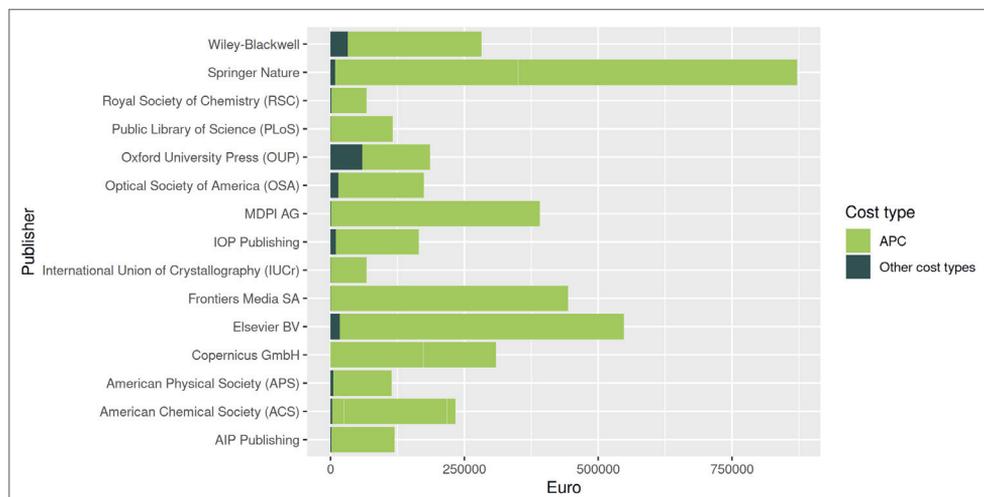


Abbildung 5: Analyse: Verteilung von APCs und Zusatzkosten nach Publisher

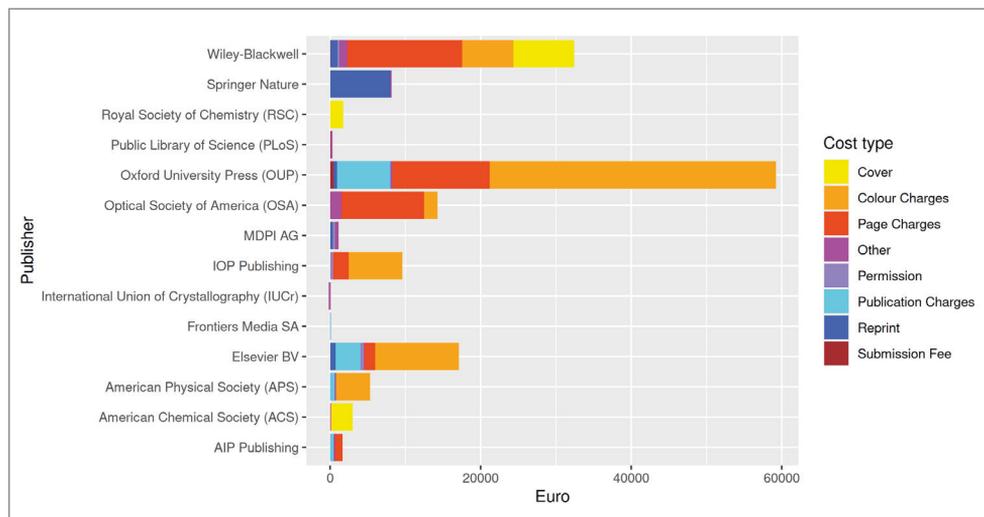


Abbildung 6: Analyse: Verteilung der Zusatzkosten nach Publishern

## Fazit und Ausblick

Insbesondere der zweite Anwendungsfall hat gezeigt, dass die Auswertung von Zusatzkosten nicht zu vernachlässigen ist und es im Rahmen des Informationsbudgets notwendig ist, diese zu erfassen, um einen umfassenden Überblick über die Ausgaben für das wissenschaftliche Publizieren zu erhalten. Die Zahlungen für den Prozess des Publizierens müssen ganzheitlich betrachtet werden, über die Publikationsfonds der Bibliotheken hinaus. Es ist eine große Herausforderung diese zusätzlichen Kosten zu ermitteln und Workflows dafür zu etablieren. Doch die Bedeutung spiegelt sich im momentanen bibliothekarischen Diskurs wider. Neben openCost gibt es aktuell noch viele weitere Projekte und Initiativen, die sich mit den Themen Informationsbudget und Kostentransparenz auseinandersetzen und mit denen das openCost-Team in engem Austausch steht; u.a. zu nennen sind die Fokusgruppe „Informationsbudget“<sup>24</sup> sowie die Projekte „Transform2Open“<sup>25</sup> und „AT2OA2“<sup>26</sup>. Allen gemeinsam ist das Bestreben, die transformativen Prozesse voranzutreiben.

OpenCost ist ein laufendes Projekt und die vorangegangenen Ausführungen haben einen Überblick über die derzeitigen Ergebnisse geliefert.<sup>27</sup> Jedoch sind bereits die weiteren Schritte in Planung bzw. in interner Vorbereitung. Die Abbildung von Transformationsverträgen und Mitgliedschaften sind Themen, die nun ausgearbeitet werden. Die aktuellen Entwicklungen des Projekts sind stets über

24 Fokusgruppe „Informationsbudget“, <<https://open-access.network/vernetzen/digitale-fokusgruppen/fokusgruppe-informationsbudget>>, Stand: 10.07.2023.

25 Transform2Open, <<https://www.transform2open.de/>>, Stand: 10.07.2023.

26 Austrian Transition to Open Access 2 (AT2OA2), <<https://www.at2oa.at/>>, Stand: 10.07.2023.

27 Projektstand vom 10.07.2023.

die Projekt-Webseite<sup>28</sup> sowie in dem openCost-Repository<sup>29</sup> auf GitHub einsehbar. Die positive Resonanz auf das Vorgehen hinsichtlich des aktuellen openCost-Metadatenchemas ist ein wesentlicher Indikator für das Projekt-Team, Ergebnisse auch weiterhin sukzessive zu veröffentlichen und begleitend Austauschformate wie Online-Diskussionen anzubieten, um zeitnah direktes Feedback aus der Community einzuholen.

### **Acknowledgement**

Diese Arbeit wurde gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – 457354095.

*Julia Bartlewski, Universitätsbibliothek Bielefeld, <https://orcid.org/0000-0001-5959-4999>*

*Christoph Broschinski, Universitätsbibliothek Bielefeld, <https://orcid.org/0000-0003-1972-7587>*

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5971>

Dieses Werk steht unter der [Lizenz Creative Commons Namensnennung 4.0 International](#).

---

28 openCost - Automatisierte, standardisierte Lieferung und offene Bereitstellung von Publikationskosten und Verlagsvereinbarungen, <<https://www.opencost.de/>>, Stand: 10.07.2023.

29 openCost Repository: GitHub – opencost-de/opencost: openCost data formats, documentation and examples, <<https://github.com/opencost-de/opencost>>, Stand: 10.07.2023.