Kommerzielles Wissen? Transparent publizieren mit openCost

Lisa-Marie Stein

1 Offene Wissenschaft braucht openCost

An das Projektteam openCost wurde in Diskussionsrunden die Frage herangetragen, warum man sich intensiv mit sogenannten Article Processing Charges[1] beschäftige, die exorbitant in die Höhe stiegen und in manchen Kreisen schlicht nicht bedienbar seien, warum etwa weniger mit Diamond-Open-Access-Prozessen.[2] Die Antwort darauf ist, dass wir die Herausforderungen im Umgang mit Publikationskosten sehen und konfrontieren. Wir sprechen uns nicht aus für ein Prinzip oder befürworten gewisse Gebührenarten, sondern fordern Transparenz in der Publikationsszene. Wir bilden ab, was uns in der Realität begegnet. Erst wenn die Karten offen auf dem Tisch liegen, können wir sie betrachten und offen über die Verhältnisse reden. Ja, sie sogar dann verstehen, verhandeln, Meinungen bilden und das Spiel in die Richtung lenken, die wir für die richtige halten. Wir können dann das Spiel gewinnen. Wir streben nach Transparenz im Publikationswesen und den damit einhergehenden Finanzströmen. Dass der Publikationsprozess Ressourcen fordert, steht dabei außer Frage. Wir wollen jedoch wissen, um welche es sich handelt und eine kommerzielle Steuerung der Wissenschaft ausschließen, in der Themen und Zielrichtung durch kommerzielle Bedürfnisse bestimmt werden. Der Zugang zu Wissen und die Veröffentlichung der eigenen Forschung sollten für jeden möglich sein, unabhängig von Fachbereich, Institution, Herkunft, finanzieller oder gesellschaftspolitischer Situation.

Die Projektinitiative openCosts ist in der Open-Access-Transformation verortet. Mit dem Begriff „Open-Access" wird der freie Zugang zu wissenschaftlicher Literatur und deren Weiterverwendung gefordert, kostenlos und technisch sowie rechtlich so barrierearm wie möglich.[3] Die „Transformation" bestrebt die vollständige Umstellung des wissenschaftlichen Publikationswesens auf Open-Access, u.a. durch die Umwandlung traditioneller Subskriptionszeitschriften in Open-Access-Zeitschriften. Die Vertreter:innen der Open-Access-Bewegung stehen für Transparenz, an der es im Publikationsprozess mangelt.

Abstract

Das wissenschaftliche Publikationswesen befindet sich in einer Transformation mit dem Ziel, offene Zugänge, faire Kriterien und Transparenz zu manifestieren. Das Projekt „openCost: automatisierte, standardisierte Lieferung und offene Bereitstellung von Publikationskosten und Verlagsvereinbarungen" stellt einen Baustein im Aufbau des angestrebten Informationsbudgets dar. openCost lässt sich als Werkzeug verstehen, um Finanzströme offenzulegen und eine automatisierte sowie standardisierte Infrastruktur im Publikationskostenmanagement einer wissenschaftlichen Einrichtung zu gewährleisten. Der vorliegende Beitrag gibt einen Überblick über die Projektinitiative, ihren methodischen Ansatz, Neuigkeiten zum openCost-Metadatenschema sowie Praxisvorteile, die sich durch die technische Schnittstelle und EZB-Erweiterung ergeben.

The scholarly publishing system is undergoing a transformation with the aim of manifesting open access, fair criteria and transparency. The project „openCost: automated, standardized delivery and open provision of publication costs and publishing agreements" represents a building block in the development of the desired information budget. openCost can be understood as a tool to disclose financial flows and to ensure an automated and standardized infrastructure in the publication cost management of a scientific institution. This article provides an overview of the project initiative, its methodological approach, news on the openCost metadata schema and the practical advantages resulting from the technical interface and EZB extension.

Thema und Prestige sind für die Wahl des Verlags entscheidend, Journal Impact Factor und H-Index werden zu kritischen Wegweisern. Wissenschaftler:innen scheinen in ihrer Publikationswahl häufig alternativlos und die Publikationskosten werden akzeptiert, sie unterliegen keinem Wettbewerb. openCost möchte einen Beitrag dazu leisten, dass an dieser Stelle Qualität als das entscheidende Kriterium im Fokus steht. Dazu ist es notwendig, die etablierten Publikationsprozesse zu hinterfragen und dabei die Finanzsituation, als einen treibenden Faktor dieser Szenerie, offenzulegen. Damit das wissenschaftsorientierte Publizieren Souveränität gewinnt, wirkt openCost daran mit, sich aus der Ohnmacht zu befreien und den

1 „Article Processing Charges" im Folgenden „APC"

2 vgl. z. B. Stein, Lisa-Marie et al.: Transparent publizieren mit openCost: Neuigkeiten zur technischen Kostendateninfrastruktur. Open-Access-Tage 2023, Berlin 2023. https://doi.org/10.3204/PUBDB-2023-06095 / https://doi.org/10.5446/66734 [3. Juli 2024].

3 vgl. Budapest Open Access Initiative, 2002. https://www.budapestopenaccessinitiative.org/translations/german-translation [3. Juli 2024]. Oder: Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen, 2003. https://openaccess.mpg.de/68053/Berliner_Erklaerung_dt_Version_07-2006.pdf [3. Juli 2024].

Abhängigkeiten wissenschaftlicher Einrichtungen, von Wissenschaftler:innen und ihren Infrastrukturen entgegenzuwirken. „Durch die Steuerung des Zugangs zu wissenschaftlicher Information über digitale Plattformen in der Hand großer Konzerne entstehen neue Märkte und problematische Effekte auf die informationelle Selbstbestimmung der Wissenschaftlerinnen und Wissenschaftler im digitalen Raum.", schreibt so auch Angela Holzer von der Deutschen Forschungsgemeinschaft.[4]

In seinen „Empfehlungen zur Transformation des wissenschaftlichen Publizierens zu Open Access"[5] befürwortet der Wissenschaftsrat 2022 die Etablierung eines Informationsbudgets an wissenschaftlichen Einrichtungen, um die Transformation strukturell voranzutreiben. „Mit Blick auf die Herausforderungen der Open-Access-Transformation dient die Schaffung eines Informationsbudgets dazu, volle Kostentransparenz herzustellen. Ziel ist es, mit diesem Instrument die Kosten für jede an der Einrichtung entstandene Publikation zu erfassen, damit eine Betrachtung der Gesamtkosten ermöglicht wird.", erläutert Heinz Pampel bereits 2019 in seiner Ausarbeitung „Auf dem Weg zum Informationsbudget […]"[6]. Bernhard Mittermaier definiert 2022 in einem Schalenmodell 26 Parameter, aus denen sich ein Informationsbudget zusammensetzt.[7] Das Informationsbudget beinhaltet sämtliche Mittel, die zur Informationsversorgung und für den Publikationsprozess anfallen. Es soll eine ganzheitliche Betrachtung ermöglichen, indem diese Kosten, unabhängig der Finanzierungsquelle, an zentraler Stelle erfasst werden. Diese Aufgabe des zentralen Kostenmonitorings und der Kostenverwaltung liegt in der Verantwortung der wissenschaftlichen Bibliotheken und erfordert Restrukturierungen der bisherigen Arbeitsprozesse sowie intensive Kommunikation und organisatorischen Zusammenschluss verschiedener Abteilungen einer wissenschaftlichen Einrichtung.

openCost formt einen konkreten Baustein zur Umsetzung dieses Konzepts des Informationsbudgets in die Praxis. U.a. entwickelt das Projektteam ein Metadatenschema zur Definition von Kostenarten und ihrer Erfassung. Bei diesem handelt es sich um ein technisches Austauschformat, das in die individuelle Infrastruktur einer Einrichtung integriert wird und den automatisierten Datenaustausch zwischen beliebigen Instanzen ermöglicht. openCost wird also zu einem Baustein des Publikationskostenmanagements. Es erfüllt den Bedarf eines automatisierten Standards, sowohl hinsichtlich optimierter Workflows in den meisten Einrichtungen als auch mit Blick auf den Informationsbedarf von Drittinstanzen, die Kostendaten erhalten möchten. Kostentransparenz erfüllt hier einen strategischen Beitrag zur kollektiven Open-Science-Bewegung: Wissenschaft und Publizierende können die notwendige Kostenkontrolle (zurück-) gewinnen, Verlagsbedingungen prüfen, informiert Position beziehen und verantwortungsbewusst Einfluss üben. Um die Open-Access-Transformation auch finanziell nachhaltig und fair zu gestalten, ist Transparenz unumgänglich.

openCost ist ein durch die Deutsche Forschungsgemeinschaft gefördertes Projekt[8] mit einer dreijährigen Laufzeit sowie regulärem Start im Dezember 2021. Die Projektpartner sind das Deutsche Elektronen-Synchrotron DESY sowie die Universitätsbibliotheken Regensburg und Bielefeld.

2 Methodischer Ansatz: Open & Community Driven

Das openCost-Metadatenschema definiert ein sogenanntes Austauschformat. Das Prinzip des Austauschs, das in diesem Zusammenhang zwar eine technische Komponente beschreibt, trifft auch auf die Arbeitsmethodik des Projekts zu: openCost verfolgt den Ansatz des offenen Diskurses. Von Beginn an wurde die Fachgemeinschaft in den Entwicklungsprozess integriert und Zwischenergebnisse in Diskussionsrunden zur Debatte gestellt.[9] So startete das Projekt zunächst mit einem Expert:innenworkshop, auf dem Anforderungen an die zu entwickelnden Formate mit internationalen Expert:innen diskutiert wurden und dessen Ergebnisse direkt in die Kostenmodellierung einflossen.[10] In einem Github-Repositorium sind Metadatenschema, Dokumentation und Kontextdiskurse einseh- und kommentierbar.[11] Der Community-Diskurs ist maßgebender Teil der Strategie. Es soll explizit kein Resultat erzielt werden, das erst nach Fertigstellung in der Praxis

4 Holzer, Angela: Die Vermessbarkeit der Wissenschaft. Digitalisierung, wissenschaftliches Publizieren, Verhaltenstracking und Wissenschaftsbewertung, in: Nicola Mößner / Klaus Erlach (Hrsg.): Kalibrierung der Wissenschaft. Auswirkungen der Digitalisierung auf die wissenschaftliche Erkenntnis, Bielefeld 2022, S. 163-181. https://doi.org/10.14361/9783839462102-007 [1. Juli 2024], S. 166.

5 Wissenschaftsrat: Empfehlungen zur Transformation des wissenschaftlichen Publizierens zu Open Access, Köln 2022. https://doi.org/10.57674/fyrc-vb61 [23. Juni 2024].

6 Pampel, Heinz: Auf dem Weg zum Informationsbudget. Zur Notwendigkeit von Monitoringverfahren für wissenschaftliche Publikationen und deren Kosten. Arbeitspapier, Potsdam 2019, S. 11. https://doi.org/10.2312/os.helmholtz.006 [23. Juni 2024].

7 vgl. Mittermaier, Bernhard: Das Informationsbudget: Konzept und Werkstattbericht, in: O-Bib. Das Offene Bibliotheksjournal Herausgeber VDB, 9 (4) (2022) S. 1-17. https://doi.org/10.5282/o-bib/5864 [23. Juni 2024], S. 6 ff.

8 DFG-Projektnummer 457354095

9 vgl. openCost-Ergebnisse der Online-Diskussionen: https://www.opencost.de/aktuelles/ergebnisse-online-diskussionen/ [21. Juni 2024].

10 vgl. Schweighofer, Bianca / Wagner, Alexander (Hrsg.): openCost on the road to publication cost transparency: proceedings, expert workshop, Hamburg, 5.–7.10.2022, Hamburg 2023. https://doi.org/10.3204/DESY-PROC-2023-01/02 [3. Juli 2024].

11 openCost-Github-Repository: https://github.com/opencost-de/opencost [21. Juni 2024].

erprobt wird, sondern eines, dessen Instrumentarium aus dem Bedarf hervorgeht und unmittelbar einsatzfähig ist. Praxisnähe ist nicht nur Impulsgeber zur Projektinitiative, sondern fortwährendes Mittel zur Überprüfung der Ergebnisse. Die Fachgemeinschaft ist eingeladen, sich an den Entwicklungsprozessen zu beteiligen. Dieser Ansatz wird auch durch die Konstellation der das Projekt betreibenden Einrichtungen verstetigt: Hier kommen heterogene Einrichtungen, zwei Universitäten und ein Forschungszentrum zusammen, in deren Publikationsstrukturen openCost zu Beginn exemplarisch erprobt wurde. OpenAPC[12] und die Elektronische Zeitschriftenbibliothek[13] stellen exemplarische Drittinstanzen für den offenen, automatisierten Datenaustausch dar. Das Forschungszentrum DESY ist darüber hinaus Mitglied des JOIN^2-Konsortiums[14], das sich aus acht verschiedenen Einrichtungen zusammensetzt, darunter mit der RWTH Aachen eine zweischichtige Universitätsbibliothek: DESYs Implementation von openCost wird allen Partnern bereitgestellt, sodass diese technisch openCost-konform aktiv werden können. In regelmäßigen Anwender:innen-Treffen werden die openCost-Projektergebnisse in die gemeinsame Repositorienstruktur integriert und diskutiert.

So profitiert openCost vom Diskurs der beteiligten Institutionen, die die ersten Projektergebnisse in ihrer eigenen Praxiswelt erproben und Erfahrungen austauschen. Es entsteht eine Dynamik zur Verifizierung des Projektmodells, besonders gewinnbringend durch die Diversität der Mitwirkenden. Das openCost-Metadatenschema entsteht so vor dem Hintergrund möglichst unterschiedlicher Anwendungsszenarien.

Vom 7. bis 9. Oktober 2024 wird ein Abschlussworkshop zur ersten Projektphase openCosts in der Universität Regensburg stattfinden, zu dem die interessierte Fachgemeinschaft eingeladen ist.

3 Schritte zur Realisierung & Resultate

Gemäß dem übergeordneten Leitmotiv der Fairness und Kostentransparenz sollen alle Publikationskosten in wissenschaftlichen Institutionen dokumentiert und belegt werden können. Zu diesem Zweck wird eine technische Infrastruktur entwickelt, mit der jene Kosten umfassend in einem standardisierten Format erfasst und mit Hilfe von standardisierten Schnittstellen offen zugänglich gemacht sowie ausgetauscht werden können.

Die konkrete praktische Realisierung dieser Ziele beinhaltet die folgenden drei Bausteine:

1. Die Ausarbeitung eines standardisierten Metadatenschemas für Publikationskosten.
2. Die Implementierung einer Harvesting-Schnittstelle zum offenen Datenaustausch.
3. Die Erweiterung des Datenmodells und Services der EZB um detaillierte Informationen zu Open-Access-Publikationskosten.

3.1 Das openCost-Metadatenschema

Im Zentrum des Projekts steht die Entwicklung des standardisierten Metadatenschemas zur strukturierten Abbildung, Erfassung und Abfrage von Publikationskosten. Im Sinne eines allumfassenden Informationsbudgets, wie es z.B. Gernot Deinzer für die Universität Regensburg beschreibt[15], soll das Schema alle Kostenarten berücksichtigen, die im Publikationsprozess anfallen. Dabei gilt es, nicht nur Open-Access-Gebühren wie die sogenannten APC, sondern alle publikationsbezogenen Gebühren, wie z.B. Farb- und Einreichungsgebühren oder durch Kofinanzierung, Mitgliedschaften und Transformationsverträge bedingte Kosten, zu erfassen. Das Metadatenschema nutzt hierzu ein normiertes Vokabular und eine strenge Schemadefinition. Formatspezifisch liegt eine XML-Repräsentation vor, die durch eine openCost-XSD-Datei[16] validiert werden kann. Um eine reibungslose Konversion in andere Formate, wie z.B. das gängige JSON, sicherzustellen, erfolgte eine attributlose Modellierung des Schemas. JSON-Serialisierungen existieren bereits für konkrete Anwendungsfälle der Projektpartner.

Ein derart standardisiertes Datenformat macht es möglich, die Daten in automatisierter Form für Kostenauswertungen zu nutzen. Dies fördert sowohl die Kostentransparenz innerhalb einer Institution als auch den Kostenvergleich zwischen verschiedenen Institutionen. Das Metadatenformat wurde in Zusammenarbeit mit nationalen und internationalen Partnern entwickelt, da es je nach Zweck der Kostenauswertung langfristig unterschiedlichen Anforderungen genügen soll. Erzielt ist eine umfangreiche, differenzierte Abbildung potentieller Eigenschaften, sodass das Schema nicht nur aktuellen, sondern auch zukünftigen Informationsbedarfen gerecht werden kann und eine nachhaltige Datenbereitstellung ermöglicht. Es soll eine verlässliche Datenbasis bieten, die nicht auf eine individu-

12 vgl. OpenAPC-Website: https://openapc.net/ [26. Juni 2024].

13 „Elektronische Zeitschriftenbibliothek" im Folgenden „EZB", Website: https://ezb.uni-regensburg.de/ [26. Juni 2024].

14 vgl. JOIN^2-Website: https://join2.de/ [21. Juni 2024] oder Stein, Lisa-Marie / Wagner, Alexander: JOIN^2 and openCost, in: Bianca Schweighofer / Alexander Wagner (Hrsg.): openCost on the road to publication cost transparency: proceedings, expert workshop, Hamburg, 5.–7.10.2022, Hamburg 2023, S. 11-30. https://doi.org/10.3204/DESY-PROC-2023-01/02 [3. Juli 2024].

15 vgl. Deinzer, Gernot / Lang, Cornelia: All in One: Informationsbudget an der Universität Regensburg. Umsetzung, Strukturen und Geschäftsprozesse, in: ZfBB Zeitschrift für Bibliothekswesen und Bibliographie 71 (2) (2024) S. 91-101. https://doi.org/10.3196/186429502471249 [3. Juli 2024]. Und: Deinzer, Gernot: Informationsbudget an der Universität Regensburg – Ein Blick über den Tellerrand, in: b.i.t.online 27 (5) (2024) S. 451-460.

16 https://github.com/opencost-de/opencost/blob/main/doc/opencost.xsd [4. Juli 2024].

```
publication [1]
    |
    |--- primary_identifier [2]
    |     |
    |     |--- doi [2.1]
    |     |
    |     | *either doi OR bibliographic_information!*
    |     |
    |     | --- bibliographic_information [2.2]
    |              |
    |              |--- Title [2.2.1]
    |              |--- Publisher [2.2.2]
    |              |--- isPartOf [2.2.3]
    |
    |--- secondary_identifiers [3]
    |     |
    |     |--- id [3.1]
    |     |     |
    |     |     |--- type = doi / handle / urn / isbn / pmid / pmc / arxiv / oai / local [3.1.1]
    |     |     |--- value [3.1.2]
    |     |
    |     |--- id
    |           |
    |           |--- [...]
    |
    |--- institution [4]
    |     |
    |     |--- id [4.1]
    |     |     |
    |     |     |--- type = ror / isni / ringold [4.1.1]
    |     |     |--- value [4.1.2]
    |     |
    |     |--- name [4.2]
    |           |
    |           |--- type = full / short [4.2.1]
    |           |--- value [4.2.2]
    |
    |--- publication_type = COAR controlled vocabulary [5]
    |
    |--- external_costsplitting = true _OR_ 1 / false _OR_ 0 [6]
```

Abbildung 1: Einzelartikel – Beschreibung Publikationsobjekt & Kostenkontext
Urheber: openCost, https://github.com/opencost-de/opencost/tree/main/doc [25. Juni 2024].
Lizenz: CC BY 4.0

```
    |--- cost_data [7]
    |     |
    |     |--- part_of_contract [7.1]
    |     |     |
    |     |     |--- primary_identifier [7.1.1]
    |     |     |     |
    |     |     |     |--- type = ESAC [7.1.1.1]
    |     |     |     |--- value [7.1.1.2]
    |     |     |
    |     |     |---    group_id [7.1.2]
    |     |
    |     |--- invoice [7.2]
    |     |     |
    |     |     |--- invoice_number [7.2.1]
    |     |     |
    |     |     |--- creditor [7.2.2]
    |     |     |
    |     |     |--- dates [7.2.3]
    |     |     |     |
    |     |     |     |--- invoice [7.2.3.1]
    |     |     |     |--- paid [7.2.3.2]
    |     |     |
    |     |     |--- amount_invoice [7.2.4]
    |     |     |     |
    |     |     |     |--- amount [7.2.4.1]
    |     |     |     |--- currency [7.2.4.2]
    |     |     |
    |     |     |--- amounts_paid [7.2.5]
    |     |           |
    |     |           |--- amount_paid [7.2.5.1]
    |     |           |
    |     |           |     |--- amount [7.2.5.1.1]
    |     |           |     |--- currency [7.2.5.1.2]
    |     |           |     |--- cost_type [7.2.5.1.3]
    |     |           |
    |     |           |--- amount_paid
    |     |                 |
    |     |                 |--- [...]
    |     |
    |     |--- invoice
    |           |
    |           |--- [...]
```

Abbildung 2: Einzelartikel – Beschreibung Kostendatenblock.
Urheber: openCost, https://github.com/opencost-de/opencost/tree/main/doc [25. Juni 2024].
Lizenz: CC BY 4.0

elle Anfrage zugeschnitten, sondern generisch abfragbar ist.

openCost präsentiert bislang jeweils ein Metadatenschema für gebührenpflichtige Einzelartikel sowie ein weiteres für Verträge. Beide Schemata setzen sich zusammen aus einer Sektion, die den Kostenkontext und das Publikationsobjekt näher beschreibt (vgl. Abb. 1 und 4), und einem Kostendatenblock, der alle publikationsbezogenen Zahlungen in Form von einem oder mehreren Rechnungselementen zusammenfasst (vgl. Abb. 2 und 5).

Die openCost-Metadaten thematisieren ausschließlich die kostenrelevanten Daten zu einer Publikation. Es ist nicht erwünscht, sonstige, an anderer Stelle bereits vorhandene Metadaten zu einer Publikation zu duplizieren. openCost liefert so einen definierten Metadatenbaustein, der in andere, individuell an einer Einrichtung gegebene Metadatenschemata eingebettet werden kann (vgl. Abb. 3).

So werden in Abb. 1 Informationen gefordert, die den Kostenkontext skizzieren: Für welche Art von Objekt wird die Transaktion fällig (`publication_type`), wer ist die zahlende Einrichtung (`institution`), die durch einen Namen und im besten Fall durch Angabe eine ROR-ID identifiziert werden kann, und liegt eine sichere Kenntnis darüber vor, dass ein externes Kostensplitting (`external_costsplitting`) stattgefunden hat. Einzig der `doi`, der als primärer Identifikator (`primary_identifier`) präferiert ist, muss an dieser Stelle als nicht kostenspezifische Information einer Publikation wiederholt werden, um den Publikationskostenbaustein eindeutig der entsprechenden Publikation zuordnen zu können (vgl. Abb. 3). Weitere Identifikatoren (`secondary_identifiers`) können optional zusätzlich angegeben werden. Wie in Abb. 2 ersichtlich, werden im Kostendatenblock (`cost_data`) die Rechnungsinformationen ausdifferenziert. Ein Rechnungselement (`invoice`) subsumiert die Rechnungsnummer (`invoice_number`), den Zahlungsempfänger (`creditor`), das Rechnungs- und/oder Bezahldatum (`dates` → `invoice`/`paid`), die Gesamtsumme auf einer Rechnung (`amount_invoice`) sowie die detaillierte Auflistung einzelner Kostenpunkte (`amounts_paid`). Eine sogenannte `amount_paid` wird dann durch Benennung des Betrags (`amount`), der Währung (`currency`) und des Kostentyps (`cost_type`) spezifiziert. Der Kostentyp enthält die Varianz herausfordernder Kostenpositionen, die es hier transparent zu machen gilt, wie z.B. `gold-oa`, `vat`, `cover charge` oder `submission fee`. Weitere zulässige Kostentypen sind durch ein feststehendes Vokabular definiert und, wie auch fortführende detaillierte Schemabeschreibungen, nachzulesen in der openCost-Dokumentation.[17]

17 openCost-Github-Repository: https://github.com/opencost-de/opencost/tree/main/doc [25. Juni 2024] und weitere Erläuterungen zum Metadatenschema für Einzelartikel bei Bartlewski, Julia / Broschinski, Christoph: Metadatenschema für die Erfassung und Übertragung von Kostendaten. Bericht aus dem „openCost"-Projekt, in: O-Bib. Das Offene Bibliotheksjournal 10(4) (2023) S. 1-13. https://doi.org/10.5282/o-bib/5971 [3. Juli 2024].

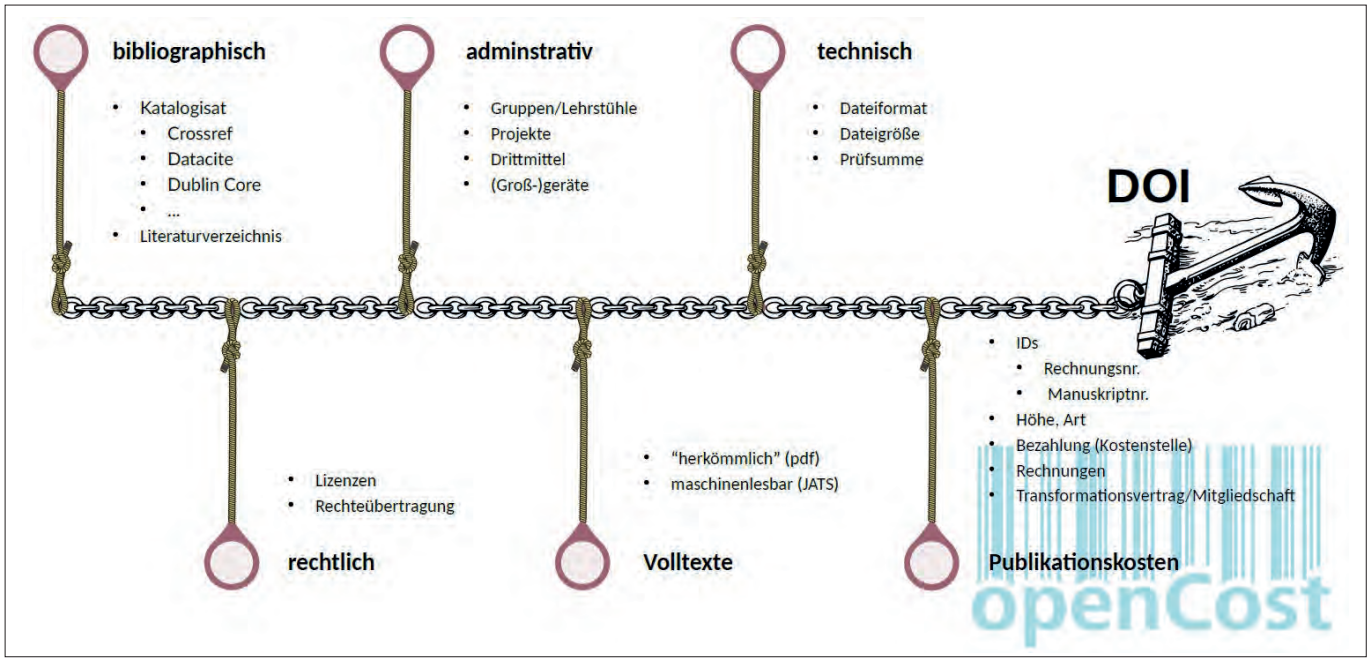

Abbildung 3: Metadaten einer Publikation samt Metadatenbaustein openCost. Urheber: Gernot Deinzer, Alexander Wagner. Lizenz: CC BY 4.0

```
contract [1]
    |
    |--- contract_name [2]
    |
    |--- institution [3]
    |     |
    |     |--- id [3.1]
    |     |     |
    |     |     |--- type = ror / isni / ringold [3.1.1]
    |     |     |--- value [3.1.2]
    |     |
    |     |--- name [3.2]
    |           |
    |           |--- type = full / short [3.2.1]
    |           |--- value [3.2.2]
    |
    |--- participation [4]
    |     |
    |     |--- from [4.1]
    |     |--- to [4.2]
    |
    |--- primary_identifier [5]
    |     |
    |     |--- type = ESAC [5.1]
    |     |--- value [5.2]
    |
    |--- secondary_identifiers [6]
    |     |
    |     |--- id [6.1]
    |     |     |
    |     |     |--- type = oai, ezb, local [6.1.1]
    |     |     |--- value [6.1.2]
    |     |
    |     |--- id
    |     |     |--- [...]
    |
```

Abbildung 4: Verträge – Beschreibung Vertrag & Kostenkontext Urheber: openCost, https://github.com/opencost-de/opencost/tree/main/doc [25. Juni 2024]. Lizenz: CC BY 4.0

Das Element `part_of_contract` (vgl. Abb. 2) verweist gegebenenfalls auf die Zugehörigkeit einer Publikation zu einem Vertrag. Über den Identifikator `ESAC`[18] wird die Publikation dann mit dem entsprechenden Vertrag verknüpft.
Insgesamt rät openCost explizit zur Verwendung der oft proklamierten persistenten Identifikatoren und sie finden sich an jeder potentiellen Position ins Schema integriert.[19]

Die Baumstrukturen der Abb. 4 und 5 visualisieren analog zu den Einzelartikeln das Metadatenschema für Verträge, das gleichermaßen für Transformationsverträge, Mitgliedschaften oder auch Subskriptionsverträge anwendbar ist. So sehr sich derartige Vereinbarungen auch inhaltlich unterscheiden, können sie doch in ihrer strukturellen Abbildung vereinheitlicht werden. Hier wurde explizit darauf geachtet, schematisch konsistent vorzugehen. Gleiches gilt in der Gegenüberstellung und für den analogen Aufbau der Schemata für Einzelartikel und Verträge, sodass bei Verträgen lediglich eine inhaltliche Erweiterung stattfindet.

18 vgl. ESAC Transformative Agreement Registry: https://esac-initiative.org/about/transformative-agreements/agreement-registry/ [25. Juni 2024].

19 vgl. Schweighofer, Bianca et al.: Enhancing Cost Transparency: The Role of Persistent Identifiers in the openCost Metadata Schema. PID fest Conference on Persistent Identifiers, Prag 2024. https://doi.org/10.3204/PUBDB-2024-04976 [4. Juli 2024].

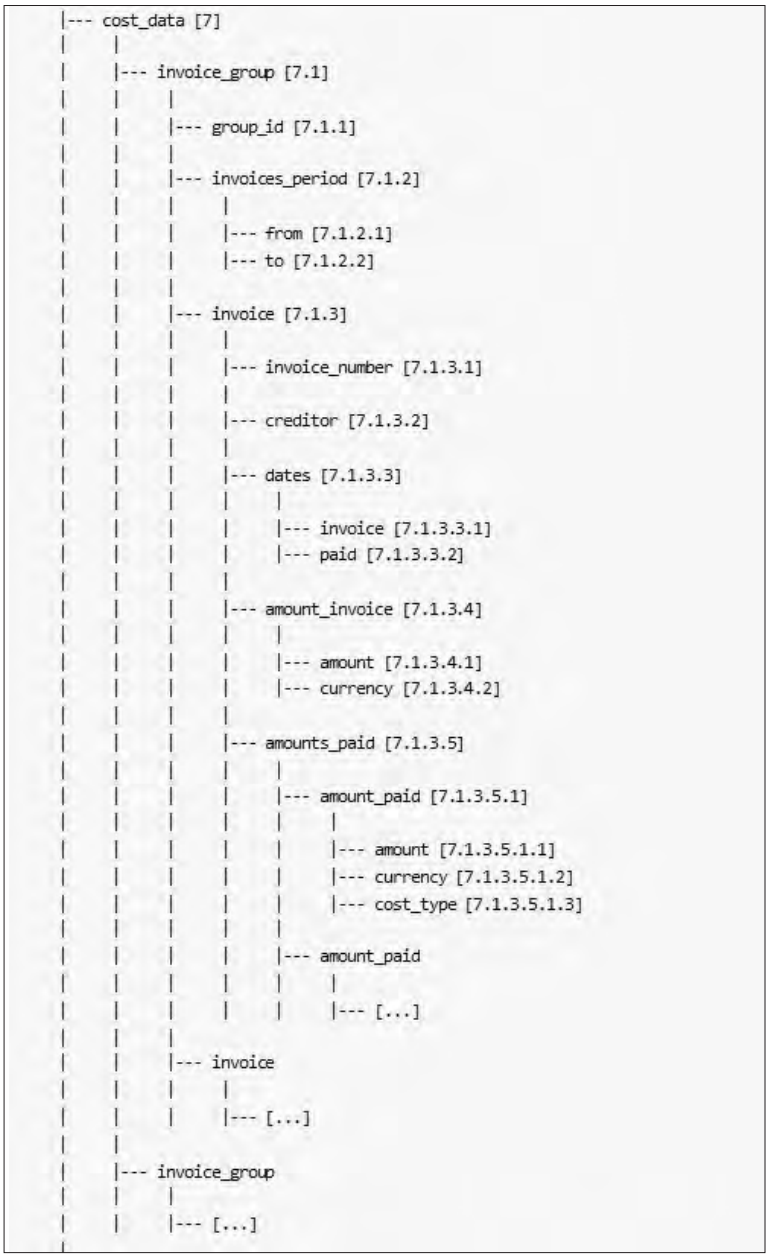

Abbildung 5: Verträge – Beschreibung Kostendatenblock
Urheber: openCost, https://github.com/opencost-de/opencost/tree/main/doc [25. Juni 2024].
Lizenz: CC BY 4.0

Bei Verträgen ist im Kostendatenblock (Abb. 5) die zusätzliche Nennung einer `group_id` notwendig, da in der Regel mehrere Rechnungen einem Vertrag zugehörig sind, die sich auf verschiedene Leistungszeiträume (`invoices_period`) beziehen. Eine solche `group_id` dient als Identifikator einer Rechnungsgruppe (`invoice_group`), die sich folglich auf eine gemeinsame `invoices_period` bezieht und aus einem oder mehreren Rechnungselementen (`invoice`) bestehen kann.

Transformationsverträge sind ein Instrument der Open-Access-Bewegung, dem aktuell besonders große Aufmerksamkeit geschenkt wird und deren Transformation, d.h. der Übergang einer subskriptionsbasierten Zeitschrift zu einem vollständigen Open-Access-Konzept, häufig noch ausstehend ist. Nicht selten fehlen jedoch für Vertragsbeteiligte abgesteckte Zeitrahmen oder weitere Orientierungen, an denen sich eine aktiv fortlaufende und verbindliche Transformation erkennen und abschätzen lässt. Zudem existiert eine Vielzahl komplexer Vertragskonzepte, die gleichsam für Beteiligte eine Herausforderung darstellen. Die Abkehr von subskriptions- und Tendenz zu publikationsbasierten oder einrichtungsfinanzierten Kostenmodellen verlangt eine Neuorientierung und -positionierung beteiligter Akteur:innen und Prozesse. Durch komplexe Kostenmodelle, Verlagsvereinbarungen und Begrifflichkeiten wird auch das zentrale Kostenmonitoring sowie auf dem Weg zum Informationsbudget Strukturen zu optimieren bzw. Strukturanpassungen zu unternehmen in wissenschaftlichen Infrastrukturen zur großen Hürde. In dieser Umbruchphase will openCost frühzeitig ansetzen, das Kostenmonitoring durch ein explizites Instrumentarium erleichtern, um Fehlentwicklungen entgegenwirken und die erwünschten Veränderungen nachhaltig gestalten zu können. Die Aufmerksamkeit ist dabei, neben großen Akteuren wie den DEAL-Konsortien im deutschen Kontext, insbesondere auch auf andere Transformationsverträge und die diversen Vertreter der internationalen Wissenschaftslandschaft zu legen.

„Die Intransparenz der Preisgestaltung und damit der Mittelverausgabung, die sich im Beschaffungswesen für wissenschaftliche Literatur in Form von Non Disclosure Agreements manifestierte, ist dabei nur eine Form der Intransparenz, die sich prinzipiell auch beim Bezug anderer Produkte ähnlich darstellt und sowohl die Abnehmer selbst betrifft als auch die Steuerung von Budgets auf übergeordneter Ebene.“, erläutert Angela Holzer und beschreibt Geheimhaltungsvereinbarungen, die zu geschwärzten Vertragszeilen führen.[20] Sowohl die Feststellung eines konkreten Preises, die Herleitung von Angeboten als auch ein Preisvergleich werden damit ausgeschlossen, eine budgetäre Planung mit Hinblick auf das Informationsbudget ist nur schwer möglich.

Das diesen Entwicklungen entgegenwirkende Konzept des Diamond-Open-Access ist entsprechend ein viel diskutierter Schwerpunkt in openCost. Auch hier gilt es, die Finanzströme standardisiert zu dokumentieren und die Schemata sind den Szenarien entsprechend anwendbar.

4.2 OAI-Schnittstelle, EZB-Erweiterung & Praxisvorteile

Sind die Publikationskostendaten also erst einmal standardisiert erfasst, profitiert die eigene Einrichtung und

20 Holzer, Angela: Die Vermessbarkeit der Wissenschaft. Digitalisierung, wissenschaftliches Publizieren, Verhaltenstracking und Wissenschaftsbewertung, in: Nicola Mößner / Klaus Erlach (Hrsg.): Kalibrierung der Wissenschaft. Auswirkungen der Digitalisierung auf die wissenschaftliche Erkenntnis, Bielefeld 2022, S. 163-181. https://doi.org/10.14361/9783839462102-007 [1. Juli 2024], S. 168.

kann sich bei der internen Informationsaufbereitung an den Inhalten und Terminologien des openCost-Metadatenschemas orientieren. Doch die Vorteile eines Standards reichen über die individuelle Nutzung hinaus: In einem zweiten Projektschwerpunkt geht es openCost um die offene Bereitstellung und den automatisierten Austausch der Kostendaten. Es finden sich diverse Anwendungsszenarien, in denen es eines solchen Austauschs bedarf und deren Prozesse durch den openCost-Standard simplifiziert werden. Die Devise lautet hier, einmalig die notwendigen Daten zu erfassen und dann auf individuelle Anfragen reagieren und sie zu jeder Zeit ohne zusätzlichen manuellen Aufwand zur Verfügung stellen zu können. Für den automatisierten Austausch schlägt openCost die etablierte OAI-PMH-Schnittstelle[21] vor, die in vielen Einrichtungen bereits vorhanden ist. Eine solche Schnittstelle macht es möglich, die Kostendaten innerhalb einer Institution, jedoch auch interinstitutionell zusammenzuführen, auszutauschen und für Auswertungen bereitzustellen.

Im Projektrahmen wird dieser Datenverkehr exemplarisch durch die Publikationsserver der Universitäten Bielefeld und Regensburg sowie des Forschungszentrums DESY – und dessen Partnerinstitutionen im JOIN^2-Verbund – im Zusammenspiel mit den Serviceinstanzen OpenAPC und EZB realisiert (vgl. Abb. 6). Die genannten Einrichtungen speichern die Kostendaten in ihren institutionellen Repositorien im openCost-Format. OpenAPC fungiert als Aggregator und sammelt via OAI-Schnittstelle die openCost-Datensätze ein, um diese auf der eigenen Plattform aggregiert und verarbeitet zur Verfügung zu stellen. Nachdem OpenAPC einst mit ihrem Ansatz zur Offenlegung der APC eine Initialzündung auf dem Weg zur Transparenz unternahm, werden sie im Rahmen openCosts ihre Datensätze um alle weiteren relevanten Kostenarten erweitern.[22]

Doch auch über OpenAPC als Harvestinginstanz hinaus können beliebige Dritte, wie z.B. Forschungsfördernde, Publikationskostendaten gezielt bei einzelnen Systemen – z.B. beim Repositorium einer Einrichtung oder unmittelbar bei OpenAPC – im openCost-Format via OAI-Schnittstelle abfragen.

Einen Beispielprozess für ein solches Anwendungsszenario stellt im deutschen Wissenschaftsraum das DFG-Pro-

21 vgl. Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH): https://www.openarchives.org/pmh/ [26. Juni 2024].

22 vgl. Bartlewski, Julia / Broschinski, Christoph: Metadatenschema für die Erfassung und Übertragung von Kostendaten. Bericht aus dem „openCost"-Projekt, in: O-Bib. Das Offene Bibliotheksjournal 10(4) (2023) S. 1-13. https://doi.org/10.5282/o-bib/5971 [3. Juli 2024].

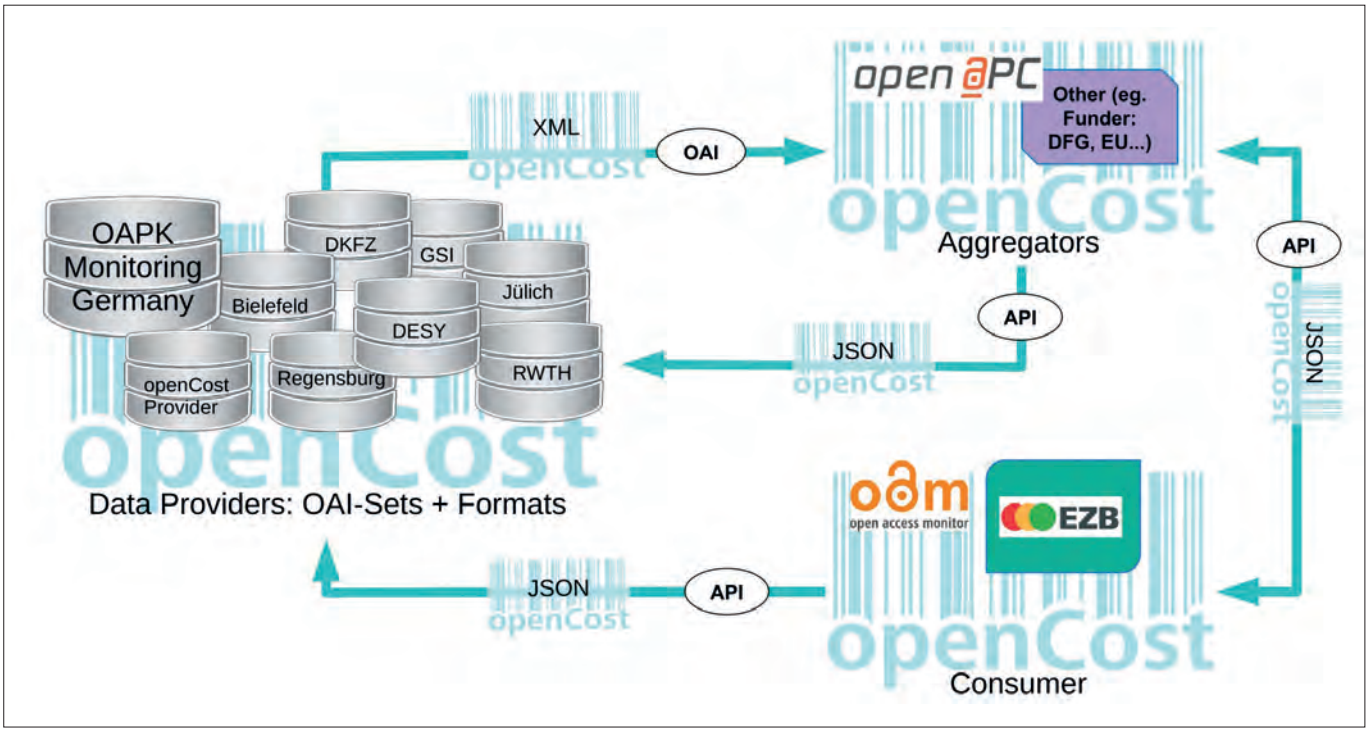

Abbildung 6: Automatischer Datenverkehr Urheber: Alexander Wagner. Lizenz: CC BY 4.0

jekt „Open-Access-Publikationskosten“[23] dar, das Open-Access-Publikationen in wissenschaftlichen Einrichtungen finanziell bezuschusst. Hierfür sind jährlich Nachweise erforderlich und die Publikationskostendaten einer Einrichtung sind an die Zentralbibliothek des Forschungszentrums Jülich zu melden.[24] Die Datenlieferung für jenes Monitoring könnte perspektivisch ebenfalls durch das openCost-Format erfolgen und würde folglich den beteiligten und openCost-produzierenden Einrichtungen den aktuell zusätzlichen Arbeitsaufwand ersparen. Kontinuierliches Harvesting wird hier auch die Erhebung eines aktuellen Zwischenstands und so eine bessere Übersicht über den tatsächlichen Finanzbedarf ermöglichen als mit einem jährlichen Verfahren möglich ist. Es findet bereits ein reger Austausch zwischen dem Forschungszentrum Jülich und dem Projekt openCost statt, der eine zunehmende Angleichung der Monitoring-Erfordernisse an die openCost-Schemainhalte erreicht. 2023 hat es einen erfolgreichen Datentransfer der an das Forschungszentrum Jülich gemeldeten Daten an openAPC via openCost-Format und -Schnittstelle gegeben und die Funktionstüchtigkeit der Gesamtdynamik wurde anhand deutschlandweiter Daten unter Beweis gestellt.

Für alle Anwendungsfälle gilt, dass die manuelle Produktion individueller Listen je nach Anforderungsbedarf durch einen automatisierten und regelmäßigen Datenverkehr ersetzt und die Aggregation der Publikationskostendaten dann direkt aus dem die Kosten erfassenden System erfolgen wird. Ein Repositorium bietet sich dafür an, da dort bereits gebündelt hochwertige Metadaten in verschiedenen Formaten vorliegen, die das Zielobjekt, die Publikation, am Ort seiner Speicherung beschreiben. Zugleich archivieren Repositorien die erworbenen Publikationen und machen sie dauerhaft weltweit verfügbar. Hier manifestiert sich das Detailwissen zu einer Publikation, während z.B. Buchhaltungssysteme in der Regel nur einen marginalen Teil abbilden und zur erforderlichen Informationsaufbereitung an Kontextwissen bedürfen. Ferner ist die OAI-Schnittstelle im Repositorium bereits vorhanden und der Datentransfer kann um Kostendaten erweitert werden.

Als inkrementelles Protokoll kann OAI die Kostendaten regelmäßig und zeitnah ‚einsammeln‘ und aktualisieren, was grundsätzlich die Vollständigkeit und Aktualität der Daten begünstigt. Statt in jährlichen oder monatlichen Abständen, können Kostendaten kontinuierlich transferiert, überprüft und gegebenenfalls aktualisiert werden. Ein Update, eine Erweiterung oder Korrektur verlangt nicht länger nach der manuellen Bearbeitung einer Liste, dessen Aktualitätsgrad stetig zu überprüfen wäre, sondern erfolgt automatisch im Rahmen der OAI-Abfrage. Ein weiterer Praxisvorteil wird mit dem dritten Projektschwerpunkt, der Erweiterung der EZB um spezielle Funktionen zur Verwaltung und Anzeige von Informationen über Open-Access-Publikationskosten bzw. die Übernahme von Publikationskosten, erzielt.

23 Förderprogramm „Open-Access-Publikationskosten” der DFG: https://www.dfg.de/de/foerderung/foerdermoeglichkeiten/programme/infrastruktur/lis/lis-foerderangebote/open-access-publikationskosten [3. Juli 2024].

24 Monitoring des DFG-Förderprogramms„Open-Access-Publikationskosten": https://www.fz-juelich.de/de/zb/open-science/open-access/monitoring-dfg-oa-publikationskosten [2. Juli 2024].

Mit der Automatisierung und Standardisierung werden folglich deutliche Ersparnisse in Arbeitsprozessen erreicht. Ein weiterer Praxisvorteil wird mit dem dritten Projektschwerpunkt, der Erweiterung der EZB um spezielle Funktionen zur Verwaltung und Anzeige von Informationen über Open-Access-Publikationskosten bzw. die Übernahme von Publikationskosten erreicht.

Dazu expandiert die Datenbasis der EZB, sodass Einrichtungen institutionsspezifische Informationen hinterlegen können: Vereinbarungen, die z.B. aus Mitgliedschaften und Transformationsverträgen resultieren, sowie Informationen zur Übernahme von Publikationskosten für bestimmte Zeitschriften. Diese Informationen können dann einrichtungsindividuell bei Auswahl der jeweiligen Zeitschrift angezeigt werden und stehen in maschinenlesbarem Format in einer API zur Nachnutzung zur Verfügung (vgl. Abb. 6).

Die EZB wird damit zur zentralen Informationsplattform für Mitarbeitende im wissenschaftlichen Publikationsprozess als auch für die publizierenden Autor:innen. In der Undurchsichtigkeit von Verlagsvereinbarungen, Kostenmodellen und Terminologien wird die EZB zum Recherchehelfer, das Recht auf Information eingefordert und praktisch realisiert. Wissenschaftler:innen können sich über Artikelpreise und die potentielle Kostenübernahme ihrer Einrichtung – technisch einfach über ein etabliertes Interface – informieren.

Wie mitunter in den Repositorien der Projektbeteiligten schon realisiert, gibt es viele denkbare Anwendungsszenarien zur Ausgestaltung des Informationsaustauschs zwischen EZB und Repositorium. In der EZB gepflegte Informationen könnten automatisiert in das eigene Repositorium oder in andere Software und Webdienste eingespeist und Workflows optimiert werden. Indem z.B. Vertragsvereinbarungen durch Verwendung der ESAC-ID systematisch in der EZB hinterlegt werden, können Vertragsinformationen für die eigene Einrichtung ins eigene System eingespielt werden. Damit werden die scheinbar simplen Fragen beantwortet, ob eine Zeitschrift einem bestimmten Vertrag unterliegt und ob die eigene Einrichtung an diesem teilnimmt, folglich welche individuellen Bedingungen und Zugriffsrechte gelten.

Abb. 6 verdeutlicht die Dynamik des automatischen Datenverkehrs auch mit Hinblick auf den Datenaustausch zwischen EZB und wissenschaftlichen Einrichtungen. Darüber hinaus wird die EZB über eine Schnittstelle mit OpenAPC verbunden, sodass die im Dienste von OpenAPC aggregrierten Publikationskostendaten auf den EZB-Detailseiten einer Zeitschrift dargestellt werden können.

4 Schlussbetrachtung

Durch die einmalige und unmittelbare Erfassung der Publikationskosten in einer wissenschaftlichen Einrichtung erheben wir die notwendige Datenbasis, statt eine Vielzahl an über die Zeit variierenden Listen zu bearbeiten und individuellen Anforderungen Dritter zu genügen. Wir profitieren von einem reduzierten Arbeitsaufwand durch den technischen Einsatz des openCost-Standards, der den automatisierten Datenaustausch ermöglicht. Um die Daten einmalig zu erfassen, muss eine individuelle interne Struktur vorhanden sein. Für den Aufbau dieser dienen die Inhalte des openCost-Metadatenschemas als Orientierung, das openCost-Austauschformat ermöglicht dann den Datentransfer. Der einrichtungsinterne und praktische Umgang mit den Projektergebnissen openCosts wird in der fortlaufenden Projektphase intensiviert. Perspektivisch soll zudem das Metadatenschema um weitere Publikationskostentypen, die sich mit der Varianz publizierter Objekte ergeben, erweitert werden.[25] openCost liefert ein Werkzeug im Aufbau einer fairen wissenschaftlichen Publikationskultur, u.a. um ihrer Kommerzialisierung vorzubeugen. „In anderen Bereichen, in denen Ergebnisse der wissenschaftlichen Forschung digital und gegebenenfalls offen zugänglich gemacht werden, z.B. im Bereich von Preprints und Forschungsdaten sowie Forschungssoftware, sollte eine solche Entwicklung hin zu einer Kommerzialisierung der wissenschaftlichen Produkte zu einem früheren Zeitpunkt unbedingt vermieden werden.“, erinnert entsprechend auch Angela Holzer.[26]

Lisa-Marie Stein (M.A.)
studierte Linguistik, Medienwissenschaft und Deutsch an der Christian-Albrechts-Universität zu Kiel. Seit 2020 ist sie im Bereich des wissenschaftlichen Publizierens tätig und ist aktuell als wissenschaftliche Mitarbeiterin in dem Projekt „openCost“ beim Forschungszentrum DESY in Hamburg beschäftigt.
ORCID: 0000-0001-7905-0462
lisa-marie.stein@desy.de

25 vgl. Stein, Lisa-Marie: The openCost project: Integrating publication cost data into institutional repositories. OAI13 – The Geneva Workshop on Innovations in Scholarly Communication, Genf 2023. https://doi.org/10.5281/ZENODO.8307954 [3. Juli 2024].

26 Holzer, Angela: Die Vermessbarkeit der Wissenschaft. Digitalisierung, wissenschaftliches Publizieren, Verhaltenstracking und Wissenschaftsbewertung, in: Nicola Mößner / Klaus Erlach (Hrsg.): Kalibrierung der Wissenschaft. Auswirkungen der Digitalisierung auf die wissenschaftliche Erkenntnis, Bielefeld 2022, S. 163-181. https://doi.org/10.14361/9783839462102-007 [1. Juli 2024], S. 168.