



Metadatenschema zum standardisierten und automatisierten Austausch von Kostendaten

Ein aktueller Bericht aus dem "openCost"-Projekt

Julia Bartlewski

Kolloquium Wissensinfrastruktur, 30.06.2023









Projekt openCost

- Automatisierte, standardisierte Lieferung und offene Bereitstellung von Publikationskosten und Verlagsvereinbarungen
- DFG-gefördert: Laufzeit 12/2021 11/2024
- Projektpartner









HINTERGRÜNDE UND ZIELSETZUNG DES PROJEKTS





Hintergründe des Projekts

- Open-Access-Transformation
 - → Umstellung des wissenschaftlichen Publikationswesens von subskriptionsbasierten Modellen zu Open Access
 - → Kostentransparenz ist notwendig, um stetig steigende Kosten zu vermeiden
- Monitoring von Publikationskosten wird zunehmend eine Voraussetzung für die Vergabe von Fördermitteln
 - → DFG-Förderprogramm "Open-Access-Publikationskosten"
 - → Übersicht über alle Publikationskosten wird angestrebt (Total Costs of Publishing)
 - → Grundlage für Budgetverwaltung
- Automatisierte und standardisierte Workflows werden benötigt!



Projektziele

- Schaffung einer technischen Infrastruktur, mit der Publikationskosten über standardisierte Schnittstellen und Formate frei zugänglich abgerufen werden können
- Nachweis möglichst aller Kosten für Publikationen wissenschaftlicher Einrichtungen

Projektschwerpunkte

- Entwicklung eines standardisierten Metadatenformats für die Erfassung und Übertragung der Daten zu Publikationskosten
- Harvesting: Basierend auf der etablierten OAI-PMH-Schnittstelle ist ein einfacher Austausch von Kostendaten möglich (intern, institutionsübergreifend, Aggregatoren)
- Erweiterung der Elektronischen Zeitschriftenbibliothek (EZB) um Informationen zu Open-Access-Publikationskosten



Speicherung von Daten zu Publikationsgebühren

- > Institutionelle Repositorien
 - → Enthalten bereits bibliographische Metadaten
 - → Idealerweise sind die zugehörigen Volltexte dort hinterlegt
 - → Verfügen über standardisierte Schnittstellen für den Austausch und Abruf der Daten



OPENCOST METADATENSCHEMA





Entwicklung eines standardisierten Metadatenformats

- Neben OA-Publikationsgebühren (APCs) sollen zusätzliche Kostendaten erfasst werden können, wie etwa:
 - Page Charges
 - Color Charges
 - Submission Fees
- Förderung der Kostentransparenz innerhalb einer Institution sowie des Kostenvergleichs zwischen verschiedenen Institutionen
 - → Informationsbudget
- Unabhängig von konkreten Diensten und Anwendungen



Einbindung der Community

openCost auf dem 8. Bibliothekskongress in Leipzig (Mai 2022)

- → Hands-on Lab "Strategie zur Erfassung der 'Total Cost of Publishing"
- → Interne Arbeitssitzung der Expert*innengruppe "Total Cost of Publishing"
 - > Arbeitsergebnisse: obib Band 9 Nr. 4 (2022), https://doi.org/10.5282/o-bib/5862

Internationaler openCost Expert*innenworkshop (Oktober 2022)

- → "openCost: the road to publication cost transparency"
 - Wissensaustausch zwischen nationalen und internationalen Expert*innen
 - Conference Proceedings, https://bib-pubdb1.desy.de/record/583603

Veröffentlichung Metadatenschema (Februar 2023)

• erste Kommentierungsphase, inkl. öffentlicher Diskussionsrunden bis 31.03.2023



Einbindung der Community (2)

OA-Tage 2022 in Bern (September 2022)

→ Vortrag: "Erweiterung der OpenAPC-Infrastruktur im Rahmen von openCost", https://doi.org/10.5281/zenodo.7802411

Workshop "Monitoring, Informationsbudget & Alma" der Landesinitiative openaccess.nrw in Essen (April 2023)

→ Vortrag: "openCost : Automatisierte, standardisierte Lieferung und offene Bereitstellung von Publikationskosten und Verlagsvereinbarungen", https://bib-pubdb1.desy.de/record/585559

1. Österreichischer Bibliothekskongress, Innsbruck (Mai 2023)

→ Vortrag: "openCost: Kostentransparenz durch Bereitstellung von Publikationskosten Überlegungen zum Datenaustausch und mögliche Anwendungen", https://bib-pubdb1.desy.de/record/585549

111.BiblioCon 2023 in Hannover (Mai 2023)

→ Vortrag: "Metadatenschema für die Erfassung und Übertragung von Kostendaten - Bericht aus dem "openCost"- Projekt", https://nbn-resolving.org/urn:nbn:de:0290-opus4-184986

Open Repositories 2023, OR 2023, Stellenbosh, South Africa (Juni 2023)

→ Vortrag: "The openCost project : Integrating publication cost data into institutional repositories", https://bib-pubdb1.desy.de/record/586562



Diskussionsergebnisse und Anforderungen

- Primäre Identifikatoren -> Bedeutung von bibliographischen Informationen
- Terminologien
- Externe Kofinanzierung
- Umgang mit Sammelrechnungen
- Umgang mit Rabatten
- Dezentrale Kosten
- Kosten ohne konkreten Publikationsbezug



Das openCost-Metadatenschema: aktuelles Zwischenergebnis

- Schema für gebührenpflichtige Zeitschriftenartikel
 - → Erweiterung des Schemas um Transformationsverträge und Mitgliedschaften im zweiten Schritt
- Zunächst im XML-Format, bietet sich speziell für den Austausch über OAI-PMH an
- GitHub Repositorium: https://github.com/opencost-de/opencost
 - → transparente Diskussion
 - → enthält neben einer tabellarischen Beschreibung des Schemas auch FAQs, eine XSD-Datei zur formalen Verifikation und einen Ordner mit Beispieldateien, die das Metadatenschema erfüllen



openCost Metadatenschema: Zahlungen für einzelne Artikel

amount_invoice	Total price as stated on the invoice. A currency attribute is mandatory.	Number	No	No	No	Net monetary value. Currency has to be a three-letter, upper case string (EUR , USD etc.) according to ISO 4217.
invoice_number	Optional invoice number	String	No	No	No	
amount_paid	An amount paid by the institution, usually corresponding to an item on the invoice. Has 2 mandatory attributes: currency and type . type identifies the cost type and has to be one of gold-oa , hybrid-oa , vat , colour charge , cover charge , page charge , permission , publication charge , reprint , submission fee , payment fee Or other .`	Number	Yes	Yes	Yes	Net monetary value. For VAT use cost type VAT. Currency has to be a three-letter, uppercase string (Eur , uso etc.) according to ISO 4217. Negative or zero values are valid to denote reimbursements or Diamond OA models.
date_paid	Date of payment. Can be a certain day, month of year, or just a year.	Pattern	No	Yes*	Yes	Format: YYYY-MM-DD, YYYY-MM or YYYY. Either this element or date_invoice is required.
date_invoice	Invoice date. Can be a certain day, month of year, or just a year.	Pattern	No	Yes*	Yes	Format: YYYY-MM-DD, YYYY-MM or YYYY . Either this element or date_paid is required.
creditor	Payment receiver as mentioned on the invoice	String	No	No	Yes	Might not be the actual publisher, but also e.g. a service provider from another publisher

https://github.com/opencost-de/opencost/tree/main/doc

```
publication
    |--- primary_identifier
             |--- doi _OR_ bibliographic_information
     --- secondary_identifiers
             |--- id type=doi
             |--- id type=local
             |--- id type=oai
    |--- institution
            |--- institution_id type=ror
            |--- institution_name type=short
    |--- type
    |--- cost_data
           |--- invoice
                  |--- amount_invoice
                  |--- dates
                        |--- date_paid
                         |--- date_invoice
                  |--- creditor
                  |--- invoice_number
                  |--- amounts_paid
                        |--- amount_paid
                        |--- amount_paid
          |--- invoice
    |--- (part of contract)
```



</opencost:publication>

openCost Metadatenschema: Beispiel

```
<opencost:publication xmlns:opencost="https://opencost.de">
  <opencost:institution>
   <opencost:institution_name type="short">desy</opencost:institution_name>
                                                                                                                       Identifier für zahlende Einrichtung
   <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>
                                                                                                                        Primärer (DOI) und sekundäre Identifier
  <opencost:secondary_identifiers>
                                                                                                                        für die Publikation
   <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>
                                                                                                                        Invoice-Block (mehrfach möglich)
     <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>
                                                                                                                   Hinweis: Möglichst auf die Angabe
     <opencost:creditor>OSA</opencost:creditor>
                                                                                                                   bibliographischer Metadaten
      <opencost:dates>
       <opencost:date_paid>2022-06-09</opencost:date_paid>
                                                                                                                   verzichten!
       <opencost:date invoice>2022-06-09
                                                                                                                   Falls erforderlich, bibliografische
     </opencost:dates>
                                                                                                                   Informationen auf andere Weise
     <opencost:invoice_number>1699976</opencost:invoice_number>
   </opencost:invoice>
                                                                                                                   hinzufügen, zum Beispiel......
  </opencost:cost data>
```



Einbettung in weitere Metadatenschemata

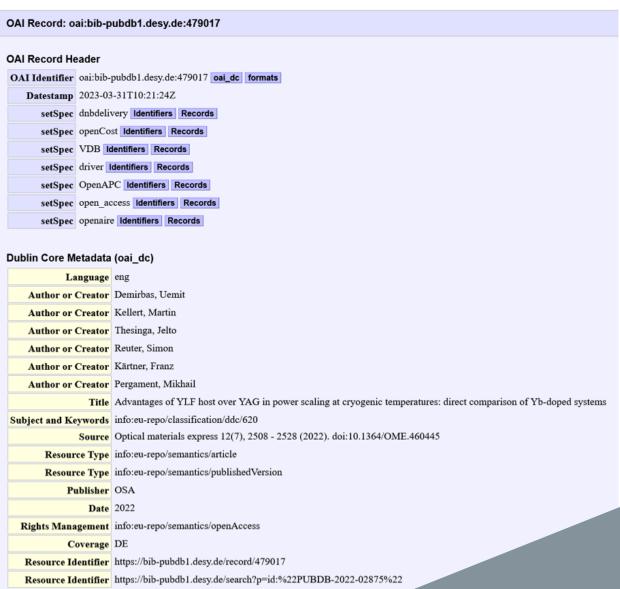
Beispiel: Erweiterung DataCite

```
<resource ..>
<identifier identifierType="DOI">10.5072/example-full</identifier>
  <creators>...</creators>
  <titles>...</titles>
  <funding references>...</funding references>
  <formats>...</formats>
  <opencost>
    <cost_data>
     <invoice>
       <amount invoice>...</amount invoice>
        <dates>
          <date paid>...</date paid>
          <date invoice>...</date invoice>
        </dates>
        <amounts paid>
          <amount_paid type="APC">...</amount_paid>
          <amount_paid type="page_charge>...</amount_paid>
        </amount paid>
     </invoice>
    </cost data>
  </opencost>
</resource>
```



Abfrage bibliografischer Metadaten mittels OAI-ID

```
<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">PUBDR-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: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:date invoice>2022-06-09
     </opencost:dates>
     <opencost:invoice_number>1699976</opencost:invoice_number>
   </opencost:invoice>
 </opencost:cost_data>
</opencost:publication>
```



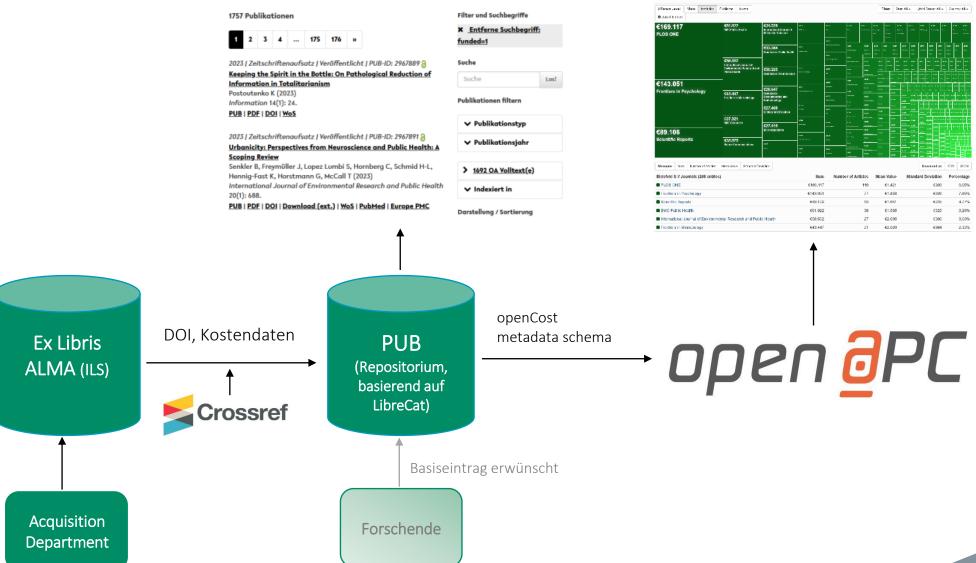


EINBINDUNG VON KOSTENDATEN IN REPOSITORIEN



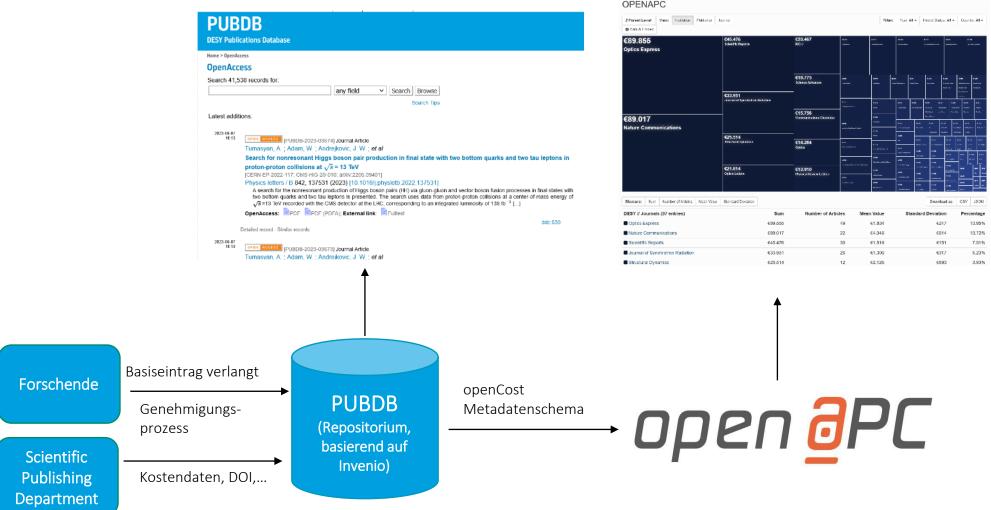


Universität Bielefeld: Kostendaten-Workflow



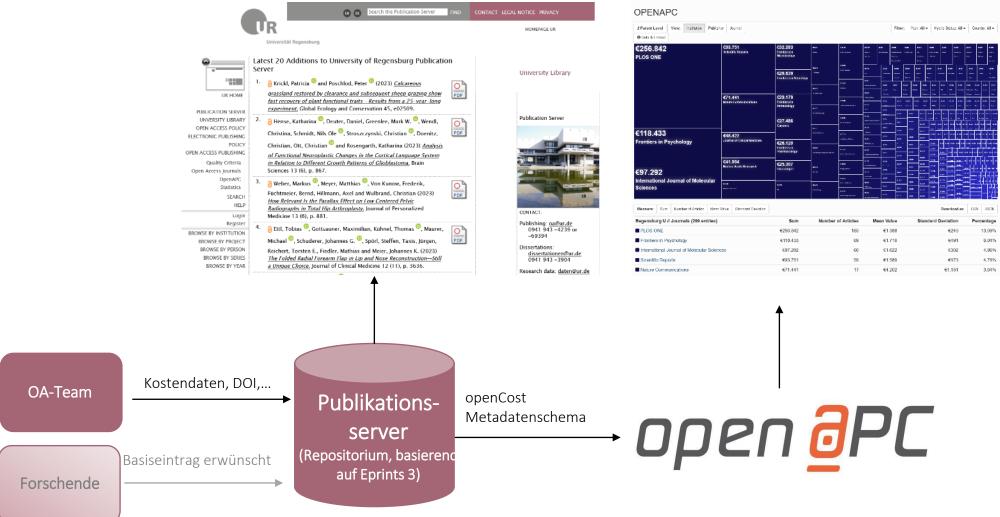


DESY: Kostendaten-Workflow





Universität Regensburg: Kostendaten-Workflow





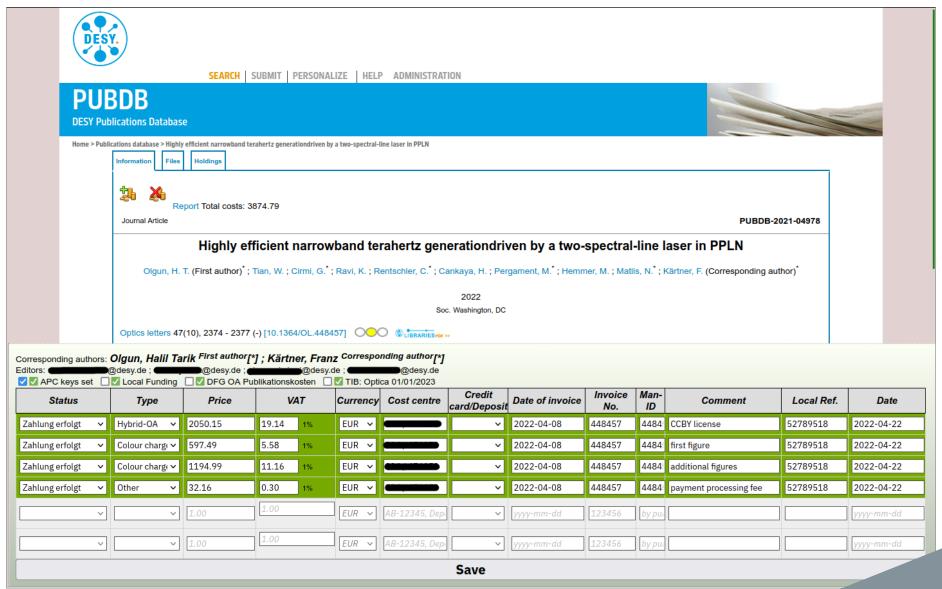
LibreCat: Eingabe von Kostendaten (Bielefeld)





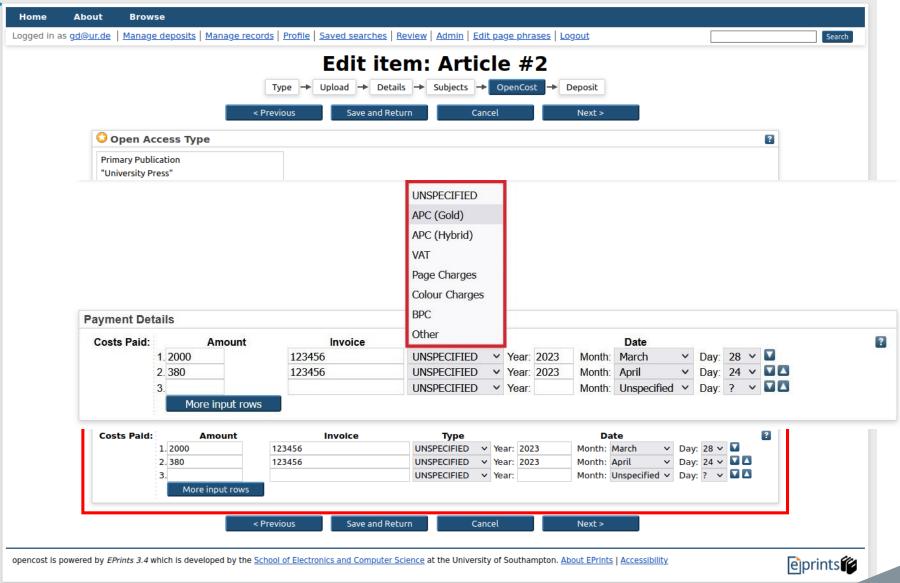


Invenio: Eingabe von Kostendaten (DESY)





EPrints: Eingabe von Kostendaten (Regensburg)





USE CASES





Anwendungsfall 1: Datenübermittlung DFG-Monitoring / OpenAPC

- Das FZ Jülich übernimmt das Kostenmonitoring im Rahmen des DFG-Programms "Open-Access-Publikationskosten"
- Teilnehmende Einrichtungen können bei Einreichung ihrer Daten festlegen, dass diese auch OpenAPC zur Verfügung gestellt werden
- Derzeitige Planung: Übermittlung der Daten unter Nutzung des OpenCost-Metadatenschemas

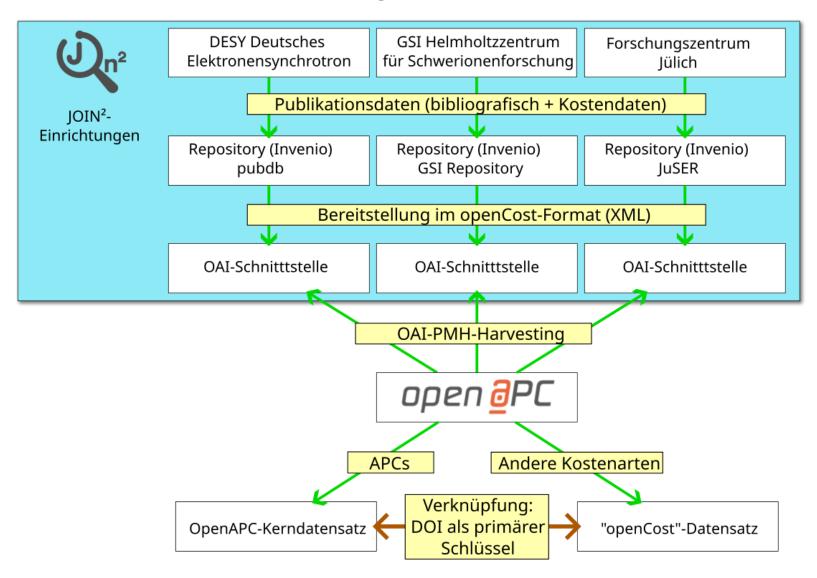


Anwendungsfall 2: Direktes Harvesting von zusätzlichen Kostendaten

- Einige Einrichtungen stellen bereits jetzt Kostendaten zu Publikationen in ihren Repositorien zur Verfügung
- In einigen Fällen kann OpenAPC diese direkt über die OAI-PMH-Schnittstelle harvesten
- Bisher: Spezialisierte Einzellösungen!
- Idee: Durch Umstellung auf das openCost-Metadatenschema können auf diese Art auch andere Gebührenarten (abseits von APCs) an OpenAPC übermittelt und dort aggregiert/ausgewertet werden
- Bereits prototypisch umgesetzt für einige Einrichtungen aus dem JOIN²-Verbund!



Übersicht: Direktes Harvesting von zusätzlichen Kostendaten





Erste Ergebnisse: Direktes Harvesting von zusätzlichen Kostendaten

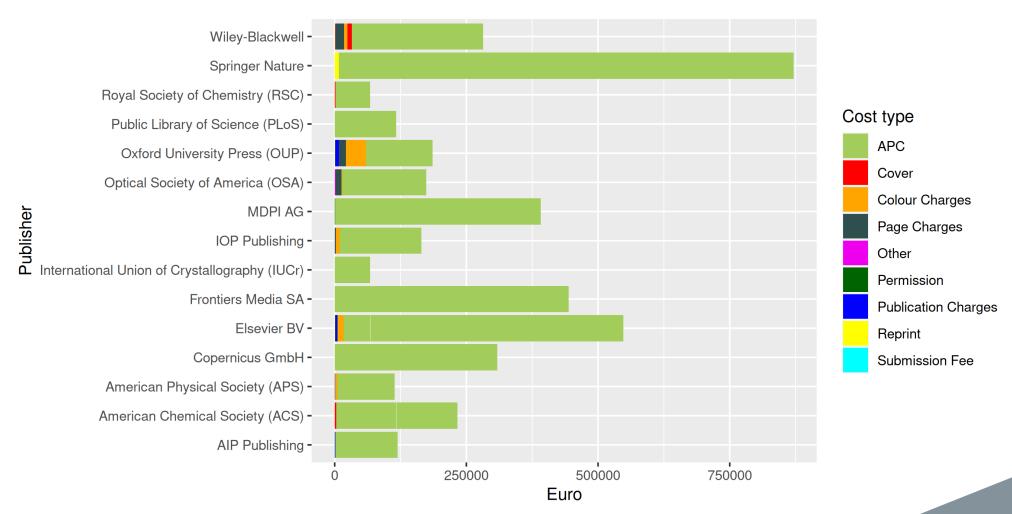
- Von den 3 JOIN²-Einrichtungen konnten insgesamt Kostendaten zu 2320 Gold-OA-Zeitschriftenartikeln geharvestet werden.
- In 268 Fällen waren zusätzlich zu einer APC weitere Kostenarten verzeichnet.

Cost type	Occurence	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



Erste Ergebnisse: Direktes Harvesting von zusätzlichen Kostendaten

Auswertung: Verteilung der zusätzlichen Kostenarten nach Verlagen







Vielen Dank!

https://opencost.de/

https://github.com/opencost-de

https://openapc.net/



