

openCost in Action

Übertragung der OAPK-Daten an OpenAPC

Christoph Broschinski, UB Bielefeld

Philipp Pollack, FZ Jülich, Zentralbibliothek

open@PC

 UNIVERSITÄT
BIELEFELD
 Universitätsbibliothek



 UR
Universität Regensburg
UNIVERSITÄTSBIBLIOTHEK

OpenCost in Kurzform

- DFG-Projekt „openCost: Automatisierte, standardisierte Lieferung und offene Bereitstellung von Publikationskosten und Verlagsvereinbarungen“
- Beteiligte: DESY Hamburg, UB Regensburg, UB Bielefeld
- Entwickelt das gleichnamige Metadatenschema „openCost“

Das openCost-Metadatenchema...

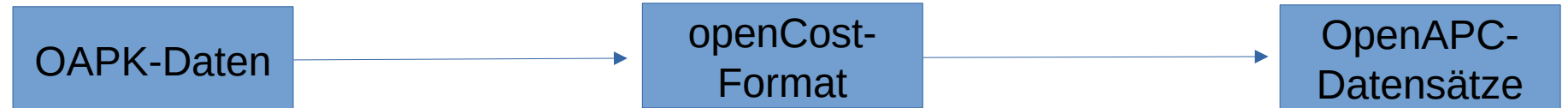
- ...ist formuliert in XML (wahlweise JSON)
- ...kann Kosten, die im Zusammenhang mit dem OA-Publizieren anfallen, darstellen (publikationsbasiert/Verträge)
- ...ist vor allem als Übertragungs-/Austauschformat gedacht

Beispiel (OAPK)

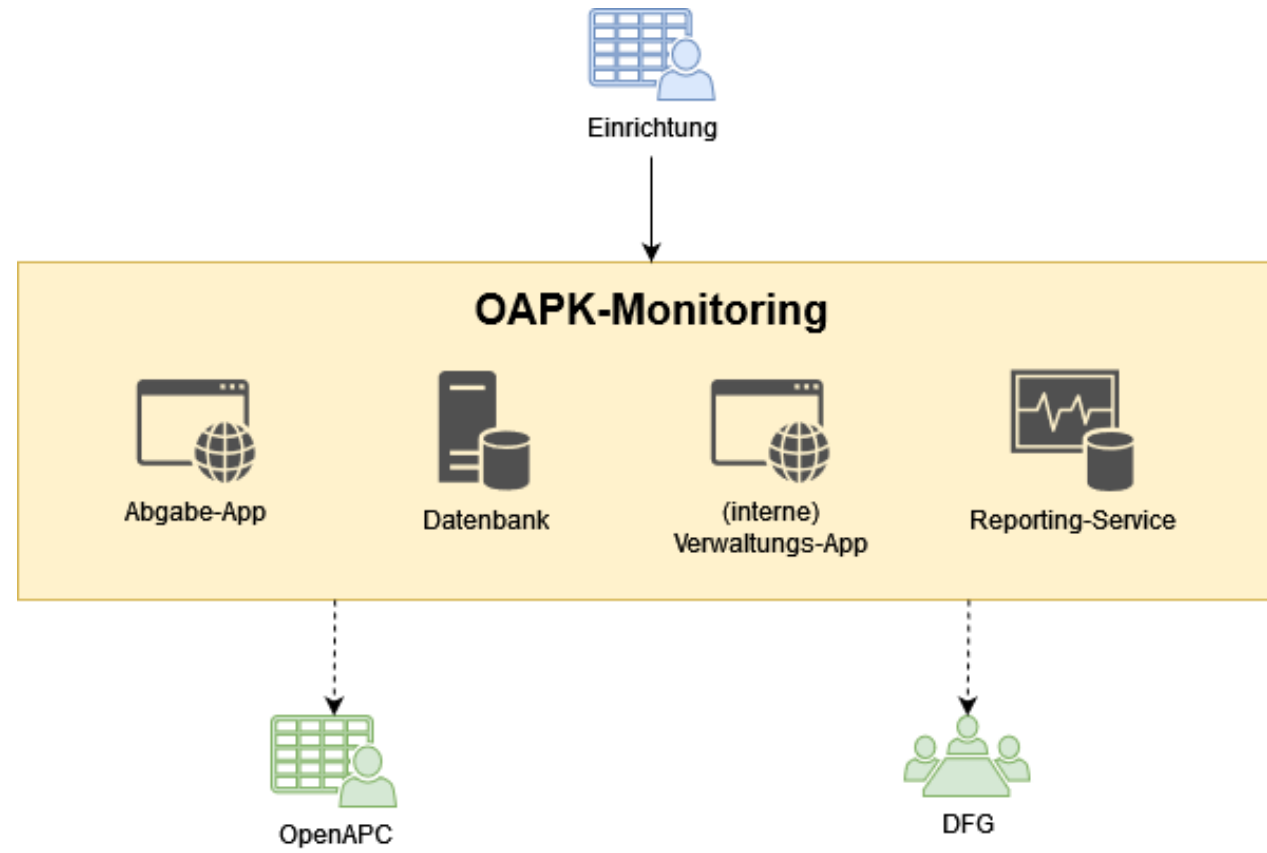
Zahlende Einrichtung	<pre><opencost:publication> <opencost:institution> <opencost:institution_id type="ror">https://ror.org/04qmmjx98</opencost:institution_id> </opencost:institution> <opencost:primary_identifier> <opencost:doi>10.5751/es-13268-270307</opencost:doi> </opencost:primary_identifier> <opencost:type>journal article</opencost:type> <opencost:cost_data> <opencost:invoice> <opencost:amounts_paid> <opencost:amount_paid currency="EUR" type="gold-oa">1056.22</opencost:amount_paid> <opencost:amount_paid currency="EUR" type="page charge">758.3</opencost:amount_paid> </opencost:amounts_paid> <opencost:dates> <opencost:date_invoice>2022</opencost:date_invoice> </opencost:dates> <opencost:amount_invoice currency="EUR">1814.52</opencost:amount_invoice> </opencost:invoice> </opencost:cost_data> </opencost:publication></pre>
Publikation	
Kostenarten	
Rechnungsbetrag gesamt	

OAPK als Anwendungsfall

- Einrichtungen konnten der Datenübertragung an OpenAPC zustimmen
- Idee: OpenCost als Übertragungsformat!
- Frage: Konvertierung?



Aufbau des OAPK-Monitoring



Einlesen der Abgaben

- Herstellen einer einheitlichen Benennung (z.B. € / Euro / EUR)
 - Konvertierung von Datentypen (z.B. Datum, Zahlen)
 - Zuordnung von ESAC-IDs
 - Regelbasierte Prüfung
- Speicherung in Datenbank

Datenübermittlung zu OpenAPC

OAPK-Metadatenchema → OpenCost-Metadatenchema

JSON → XML

Datenübermittlung zu OpenAPC - Verträge

- Ein Element pro Vertrag und Einrichtung
- Kosten: Gruppierung nach Rechnungsjahr und Lizenzjahr
 - InvoiceId: {Ror}_{Esac}_{Rechnungsjahr}_{Lizenzjahr}
 - AmountsPaid: Ein Eintrag pro Position (Netto) zzgl. je ein Eintrag für die Steuern (sofern Steuersatz >0%)
 - InvoicePeriod: Lizenzjahr
 - DateInvoice: Rechnungsjahr

Datenübermittlung zu OpenAPC - Publikationen

- Gruppierung nach DOI, Dokumenttyp, Einrichtung, Vertrag, Rechnungsjahr und Publikationsjahr
- Kosten: Für jedes Kostenelement
 - Falls zu Vertrag gehörig, finden der zugehörigen InvoiceId: Vertragsname gleich, Einrichtung gleich, DateInvoice = Rechnungsjahr, InvoicePeriod = Publikationsjahr
 - AmountsPaid: Ein Eintrag pro Position (Brutto) mit dazugehöriger Kostenart

Umwandlung von DEAL-Kostendaten in OpenAPC

- In openCost werden DEAL-Kostendaten **vertragsbasiert** dargestellt (sog. *contract*-Elemente)
- OpenAPC verwendet hingegen grundsätzlich **publikationsbasierte** Kosten
- Bereits seit Längerem verfolgt OpenAPC den Ansatz der *Equivalent APC*: Die jährlichen Kosten werden auf Artikel in hybriden DEAL-Zeitschriften umgelegt
- Formel:
$$\text{Kosten pro Artikel} = \text{Jährliche DEAL-Kosten (PABA)} / \text{Anzahl Artikel}$$
- Wie werden die entsprechenden Daten umgewandelt?

Beispiel: Wiley-DEAL, Universität Regensburg

Situation: Die übermittelte openCost-XML-Datei enthält 370 *contract*-Elemente

```
<opencost:contract>
  <opencost:contract_name>Wiley (DEAL) 2019-2023</opencost:contract_name>
  <opencost:institution>
    <opencost:institution_id type="ror">https://ror.org/01eezs655</opencost:institution_id>
  </opencost:institution>
  <opencost:primary_identifizier type="ESAC">wiley2019deal</opencost:primary_identifizier>
  <opencost:cost_data>
    <opencost:invoice>
      <opencost:amounts_paid>
        <opencost:amount_paid currency="EUR" type="publishing">136449.86</opencost:amount_paid>
        <opencost:amount_paid currency="EUR" type="vat">25925.47</opencost:amount_paid>
        <opencost:amount_paid currency="EUR" type="read">131098.89</opencost:amount_paid>
        <opencost:amount_paid currency="EUR" type="vat">9176.92</opencost:amount_paid>
      </opencost:amounts_paid>
      <opencost:dates>
        <opencost:date_invoice>2021</opencost:date_invoice>
      </opencost:dates>
      <opencost:amount_invoice currency="EUR">302651.14</opencost:amount_invoice>
      <opencost:invoice_id>01eezs655_wiley2019deal_2021_2022</opencost:invoice_id>
      <opencost:invoice_period>
        <opencost:from>2022</opencost:from>
        <opencost:to>2022</opencost:to>
      </opencost:invoice_period>
    </opencost:invoice>
  </opencost:cost_data>
</opencost:contract>
```

Schritt 1: Identifizieren von DEAL-Verträgen (ESAC-ID)

Schritt 2: Auslesen der Kosten

Schritt 3: Auslesen der *invoice_id*, um verknüpfte Publikationen zu finden

Beispiel: Wiley-DEAL, Universität Regensburg (2)

Situation: Die übermittelte openCost-XML-Datei enthält 15.720 *publication*-Elemente

```
<opencost:publication>
  <opencost:institution>
    <opencost:institution_id type="ror">https://ror.org/01eezs655</opencost:institution_id>
  </opencost:institution>
  <opencost:primary_identifier>
    <opencost:doi>10.1111/plb.13385</opencost:doi>
  </opencost:primary_identifier>
  <opencost:type>journal article</opencost:type>
  <opencost:cost_data>
    <opencost:part_of_contract>
      <opencost:primary_identifier type="ESAC">wiley2019deal</opencost:primary_identifier>
      <opencost:invoice_id>01eezs655_wiley2019deal_2021_2022</opencost:invoice_id>
    </opencost:part_of_contract>
  </opencost:cost_data>
</opencost:publication>
```



institution	period	euro	doi	is_hybrid	type
Regensburg U	2022	4145.91	10.1111/plb.13385	TRUE	journal article
Regensburg U	2022	4145.91	10.1111/obr.13422	TRUE	journal article
Regensburg U	2022	4145.91	10.1002/ccd.30046	TRUE	journal article
Regensburg U	2022	4145.91	10.1111/sji.13146	TRUE	journal article
Regensburg U	2022	4145.91	10.1002/zaac.202100378	TRUE	journal article
Regensburg U	2022	4145.91	10.1111/php.13594	TRUE	journal article
Regensburg U	2022	4145.91	10.1111/eea.13149	TRUE	journal article

Schritt 4: Suchen aller *publication*-Elemente, die die zuvor ausgelesenen *invoice_id* tragen

Schritt 5: Bestimmen der Institution

Schritt 6: Zählen der gefundenen *Publication*-Elemente (hier: 73)

Schritt 7: Berechnung der *Equivalent APC*:

$$302.651,14\text{€} / 73 = 4.145,91\text{€}$$

Schritt 8: Umformung aller Metadaten in das bekannte OpenAPC-Tabellenformat

Probleme / Herausforderungen

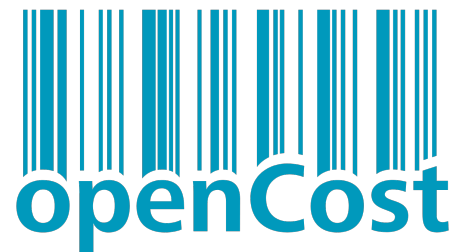
- Beschriebener Workflow konnte in diesem Jahr noch nicht voll umgesetzt werden aufgrund Abweichung OAPK-Daten / DEAL-PABAs → individuelle Nachmeldungen der Artikellisten durch Einrichtungen an OpenAPC
- Jahreszuordnung in openCost-Format nicht eindeutig (Rechnungsjahr/Lizenzjahr?)
- Übermittlung von Nachzahlungen/Rückerstattungen noch unklar
- OpenCost-Format noch nicht finalisiert, Anpassungen bei der nächsten Lieferung nötig

Fazit / *lessons learned*

- Die OAPK-Daten konnten durch Jülich erfolgreich in das openCost-Format überführt und an OpenAPC übermittelt werden
- Auf OpenAPC-Seite waren größere Transformationsschritte nötig – dies ist aber im openCost-Kontext so gedacht (Aggregator ist in der Pflicht, nicht das Format!)

Ausblick

- Workflows / Skripte beidseitig etabliert, zukünftige Übermittlungen sollten daher reibungslos ablaufen
- DEAL-Daten zukünftig auf Grundlage der PABAs → Datenqualität gesichert
- In Zukunft auch Nachnutzung anderer Kostenarten in OAPK durch OpenAPC (page charges, color charges...)



Vielen Dank!

www.opencost.de