Repositories as a Corner Stone of Publication Cost Transparency

XML, OAI-PMH and all that

Alexander Wagner

Göteborg, 04.06.2024

HELMHOLTZ

Overview

> Who and what is openCost

> Cost Bearing Entities

> Cost Capture and Storage

> The openCost Format
- Individual Payments
- Electronic Publishing Contracts

> Outlook

Who is openCost?

Repository: **EPub** → web services, journal & contract data

Repository: **Pub** → web services OpenCAPC data

Repository: **PubDB** → web services consumer

≈ 610.000 records, 98.000 OpenAccess, > 23.000 staff, > 6.000 visiting scientists
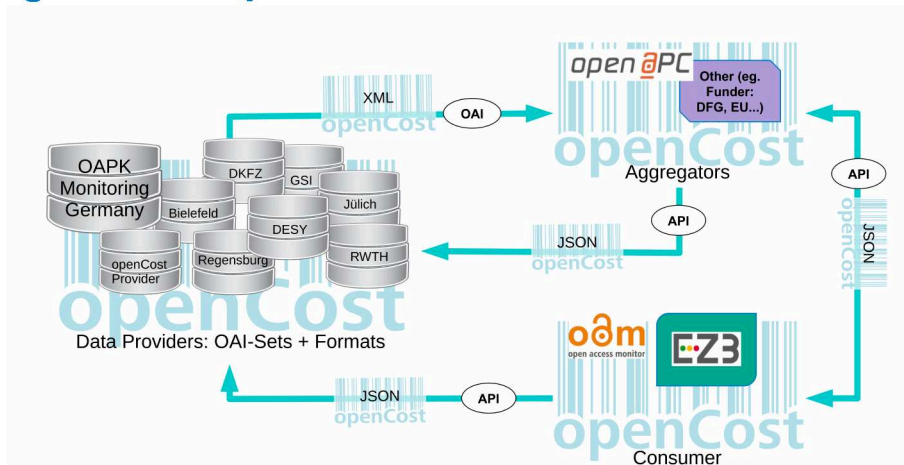
Its all about money…

TL;DR

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

Therefore, we are developing a standardized metadata schema to record, retrieve and map all publication costs of a scientific institution in a structured form. In addition to OA publication charges (APCs), it also includes costs from transformative agreements, memberships, etc. We are proposing the established OAI-PMH interface for an automatic exchange.

https://opencost.de/en | https://github.com/opencost-de/opencost

The Big Picture: openCost Data Flow

Timely, regular provision of cost data is imperative for subsequent functions!

Cost Bearing Entities

Costs are stored with the cost bearing entity

Publications

> **A**PC: Articles
> **B**PC: Books
> **C**PC: Chapters
> **D**PC: Data sets
> …

Classical items in Repositories

Cost Bearing Entities

Costs are stored with the cost bearing entity

Publications

> APC: Articles

> BPC: Books

> CPC: Chapters

> DPC: Data sets

> …

EPC: Electronic Publishing Contracts

(cf. Deinzer et al. 06/June/2024: 09:00 - 10:30 - Drottningporten 1)

> "Transformative" Agreements

> Memberships

> Subscribe2Open

> Diamond Open Access

Classical items in Repositories

New to repositories

Cost Capture and Storage

openCost currently defines an *exchange* format!

Internal formats require extensions for workflow etc.

> Model the invoice
> Use a simple, fast, straight forward UI

Cost Capture and Storage

openCost currently defines an *exchange* format!

Internal formats require extensions for workflow etc.

> Model the invoice
> Use a simple, fast, straight forward UI
> Cost types can appear more than once
> Costs have a sign: $\pm$
> Aggregators sum identical cost types
> Mind the currencies
> Store price in original currency
> Prices are net (VAT is a cost type)

Type	Price	Currency
APC	2675.00 **100.00 %**	USD
APC	-700.00	EUR
APC	700.00	EUR

Example: USD invoiced, central cofunding

openCost: Who and What

```
|--- * primary_identifier      [2]
|       |--- doi               [2.1]
|
|--- secondary_identifiers     [3]
|      |--- id                 [3.1]
|      |      |--- type        [3.1.1]
|      |      |--- value       [3.1.2]
|      |--- id
|             |--- [...]
|
|--- * institution             [4]
|      |--- * id               [4.1]
|      |      |--- type        [4.1.1]
|      |      |--- value       [4.1.2]
|      |
|      |--- name               [4.2]
|             |--- type        [4.2.1]
|             |--- value       [4.2.2]
|--- * publication_type        [5]
|--- external_costsplitting    [6]
```

opencost.xsd

```xml
<oc:publication xmlns:oc="opencost.xsd">
<oc:primary_identifier>
  <oc:doi>10.1088/1475-7516/2024/03/007</oc:doi>
<oc:/primary_identifier>
<oc:secondary_identifiers>
  <oc:id>
    <oc:type>oai</oc:type>
    <oc:value>oai:bib-pubdb1.desy.de:604831</oc:value>
  <oc:/id>
  <oc:id>
    <oc:type>arxiv</oc:type>
    <oc:value>2312.12360</oc:value>
  <oc:/id>
  <oc:id>
    <oc:type>doi</oc:type>
    <oc:value>10.3204/PUBDB-2024-01241</oc:value>
  <oc:/id>
<oc:/secondary_identifiers>
<oc:institution>
  <oc:id>
    <oc:type>ror</oc:type>
    <oc:value>https://ror.org/01js2sh04</oc:value>
  <oc:/id>
  <oc:name>
    <oc:type>short</oc:type>
    <oc:value>desy</oc:value>
  <oc:/name>
<oc:/institution>
<oc:publication_type>journal article<oc:/publication_type>
```

openCost: Invoices

```
|- * cost_data                        [7]
|   |
|   |--- * 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
|        |--- [...]
```

```xml
<oc:cost_data>
  <oc:invoice>
    <oc:amounts_paid>
      <oc:amount_paid>
        <oc:amount>0.00</oc:amount>
        <oc:currency>EUR</oc:currency>
        <oc:cost_type>hybrid-oa</oc:cost_type>
      </oc:amount_paid>
    </oc:amounts_paid>
    <oc:creditor>IOP</oc:creditor>
    <oc:dates>
      <oc:paid>2024-04-19</oc:paid>
      <oc:invoice>2024-04-19</oc:invoice>
    </oc:dates>
    <oc:invoice_number>1</oc:invoice_number>
  </oc:invoice>
</oc:cost_data>
```

> Add as many `invoices` as relevant [1..n]
> Add as many `amounts_paid` as `cost_types` [1..n]
> `creditor` is not always the publisher
> Vocabularies: `cost_type`, `currency`

EPCs: Electronic Publishing Contracts

Publications

```
|--- cost_data                      [7]
    |--- part_of_contract           [7.1]
    |   |--- primary_identifier     [7.1.1]
    |   |   |--- type               [7.1.1.1]
    |   |   |--- value              [7.1.1.2]
    |   |--- group_id               [7.1.2]
```

Builds link to contract and payment period.

> `primary_identfier` → contract

> `group_id` → relevant time span

(=the group of relevant invoices)

> Cost bearing entity: contract OR discounted due to a contract

> Contracts span a time period

(payment for n publications; n unknown beforehand)

> Requires PID for the contract

(ideally provided by some well known registry)

> `group_id` refers to a group of invoices

(e. g. a uid, `<ror>-<contractid>-<year>`)

> Publications may still bear costs

(Costs not covered by the contract)

EPC Structure

```
|-- contract_name        DEAL
|-- institution          https://ror.org/01js2sh04
|-- participation        2020-01-01 - 2023-12-31
|-- primary_identifier   ESAC:deal2020
|
|-- cost_data
|   |-- invoice_group            desy-deal2020-2020
|   |   |-- invoices_period      2020-01 - 2020-12
|   |   |-- invoice
|   |   |   |-- invoice_number   2020-01
|   |   |   |-- amounts_paid
|   |   |       |-- cost_type    publish
|   |   |       |-- cost_type    read
|   |   |       |-- cost_type    vat
|   |   |-- invoice
|   |   |   |-- invoice_number   2020-02
|   |   |   |-- amounts_paid
|   |   |       |-- cost_type    publish
|   |   |       |-- cost_type    read
|   |   |       |-- cost_type    vat
|   |
|   |-- invoice_group            desy-deal2020-2021
|   |   |-- invoices_period      2021-01 - 2021-12
|   |   |-- invoice
|   |   |   |-- invoice_number   2021-01
|   |   |   |-- amounts_paid
|   |   |       |-- cost_type    publish
|   |   |       |-- cost_type    read
|   |   |       |-- cost_type    vat
```

> Number of invoices per year varies

> Publications in $202x \rightarrow$ `desy-deal2020-202x`

> Articles may be part of a group, but
> - bear additional costs (e. g. page charges etc.)
> - are paid individually (e. g. with some "discount")

> Total costs in $202x$:
>   all `amounts_paid` for $202x$
> + additional costs
> + individually paid articles

Proper association is important!

Staff needs a helping hand…

OAI Setup

> Metadata Format: `openCost`
> OAI-Set: `openCost`

The set should expose all records with cost data

> individual articles
> EPCs (aka contract objects)
> articles linked to EPCs

WIP: Additional Reporting Requirements

> openCost focuses on cost data
> Stand-alone usage via PID (doi)
> openCost is embeddable…
> …and meant to be embedded

WIP: Additional Reporting Requirements

> openCost focuses on cost data
> Stand-alone usage via PID (doi)
> openCost is embeddable…
> …and meant to be embedded

DFG Open Access monitoring:
> Funding information
> Subject classification

```
datacite:resource
|
|--- opencost:publication
|    |--- opencost:primary_identifier
|    |    |--- doi
|    |--- opencost:institution
|    |    |--- [...]
|    |--- opencost:publication_type
|    |--- opencost:cost_data
|         |--- invoice
|              |--- [...]
|
|--- datacite:fundingReferences
|    |--- datacite:fundingReference
|    |    |--- datacite:awardTitle
|    |    |--- datacite:funderidentifier
|    |    |--- datacite:fundername
|    |    |--- datacite:awardNumber
|    |
|    |--- datacite:fundingReference
|    |    |--- [...]
|    |
|    |--- datacite:subjects
|    |    |--- datacite:subject
|    |    |--- datacite:subject
```

Outlook: openCost 2

> Define an internal storage format
> Hook up with ILS systems
> Interactive openCost (e. g. JSON-serialization)
> Enhance coverage
  - Diamond Models
  - Conference proceedings
  - Books / Book Chapters
  - Data / Software Publications
> Enhance EZB to a contract registry

Join the discussion!
 https://github.com/opencost-de/opencost

Thank you!

For more on openCost see also

> Schweighofer et al.
> *The OpenCost project - Using repositories to achieve transparency in the transformation process to open access as part of the science economy*
> Wed, June, 5th: 09:00 - 10:30 · Drottningporten 3

> Deinzer et al.
> *Transparency of payments for Open Access − Extending the metadata set with contract and payment details*
> Thu, June, 6th: 09:00 - 10:30 · Drottningporten 1

> **The Road to Publication Cost Transparency - The Next Stage**
> **October 7th − 9th, Regensburg (Germany)**

Contact

Deutsches Elektronen-
Synchrotron DESY

www.desy.de

Alexander Wagner
0000−0001−9846−5516
Central Library
alexander.wagner@desy.de
+49−40−8998−1758
10.3204/PUBDB-2024-01619

Typeset by luaLaTeX