

The openCost project

Integrating publication cost data into institutional repositories

Open Repositories 2023

13/June/2023 | Julia Bartlewski (Bielefeld University)

Christoph Broschinski, Gernot Deinzer, Dirk Pieper, Alexander Wagner



openCost

- Automated, standardized delivery and open provision of publication costs and publishing agreements
- Funded by the German Research Foundation (DFG)
 - Term: 12/2021 - 11/2024
- Project partners:



STARTING POINT AND PROJECT OBJECTIVES



Context of the project

- Open Access Transformation
 - Shifting system of scientific publishing from subscription based models to Open Access
 - Cost transparency is imperative in order to avoid ever increasing costs
 - Monitoring of publication-related costs is becoming a prerequisite for funding
 - Overview of all publication costs required (Total Costs of Publishing)
 - Foundation for budget administration
- Automated and standardized workflows are needed!

Project goals

- Creating a technical infrastructure that allows to comprehensively record all costs involved in scientific publishing and subsequently make them freely accessible by means of standardized interfaces
- Enabling cost transparency on an institutional, national, and international level

Project focus

- Standardized metadata schema: This schema will allow to collect and expose all costs involved in scientific publishing
- Harvesting: Building upon the well-established OAI-PMH interface allows easy exchange of locally collected cost data with aggregator platforms

Where to store data on publication fees?

- Institutional Repositories
 - Already store bibliographic metadata
 - Ideally hold full texts
 - Feature standardized interfaces for access and sharing

OPENCOST METADATA SCHEMA



Developing a standardized metadata schema

- In addition to OA publication charges (APCs), it should also include costs from transformative agreements, memberships, non-OA (e.g. page or colour charges) costs as well as costs for other processing fees (like costs for bank transfer, credit cards etc.)
 - Enabling individual institutions to obtain an overview of all payments
 - Enabling comparison between different institutions
- First step: Metadata schema for fee-based, Open Access journal articles
- Proposing an XML serialization of the openCost format → Using the OAI-PMH protocol for data exchange



openCost metadata schema: Payments for individual articles

<code>amount_invoice</code>	Total price as stated on the invoice. A <code>currency</code> 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.
<code>invoice_number</code>	Optional invoice number	String	No	No	No	
<code>amount_paid</code>	An amount paid by the institution, usually corresponding to an item on the invoice. Has 2 mandatory attributes: <code>currency</code> and <code>type</code> . <code>type</code> identifies the cost type and has to be one of <code>gold-oa</code> , <code>hybrid-oa</code> , <code>vat</code> , <code>colour charge</code> , <code>cover charge</code> , <code>page charge</code> , <code>permission</code> , <code>publication charge</code> , <code>reprint</code> , <code>submission fee</code> , <code>payment fee</code> or <code>other</code> .	Number	Yes	Yes	Yes	Net monetary value. For VAT use cost type VAT. Currency has to be a three-letter, uppercase string (EUR , USD etc.) according to ISO 4217. Negative or zero values are valid to denote reimbursements or Diamond OA models.
<code>date_paid</code>	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 <code>date_invoice</code> is required.
<code>date_invoice</code>	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 <code>date_paid</code> is required.
<code>creditor</code>	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

```
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)
```

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



openCost metadata schema: Sample record

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



Identifier for the institution which *pays for the publication*



Primary and secondary identifier for the publication



Invoice block (can occur more than once)

- Important: Abstain from including bibliographic metadata as far as possible!
- If needed, bibliographic information should be added in other ways, for example...

Integration into other metadata schemas

Example: DataCite extension

```
<resource ..>
<identifier identifierType="DOI">10.5072/example-full</identifier>
  <creators>...</creators>
  <titles>...</titles>
  <funding referneces>...</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>
```

Getting bibliographic metadata via 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">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:date_invoice>2022-06-09</opencost:date_invoice>
      </opencost:dates>
      <opencost:invoice_number>1699976</opencost:invoice_number>
    </opencost:invoice>
  </opencost:cost_data>
</opencost:publication>
```



OAI Record: oai:bib-pubdb1.desy.de:479017

OAI Record Header

OAI Identifier	oai:bib-pubdb1.desy.de:479017	oai_dc	formats
Datestamp	2023-03-31T10:21:24Z		
setSpec	dnbdelivery	Identifiers	Records
setSpec	openCost	Identifiers	Records
setSpec	VDB	Identifiers	Records
setSpec	driver	Identifiers	Records
setSpec	OpenAPC	Identifiers	Records
setSpec	open_access	Identifiers	Records
setSpec	openaire	Identifiers	Records

Dublin Core Metadata (oai_dc)

Language	eng
Author or Creator	Demirbas, Uemit
Author or Creator	Kellert, Martin
Author or Creator	Thesinga, Jelto
Author or Creator	Reuter, Simon
Author or Creator	Kärtner, Franz
Author or Creator	Pergament, Mikhail
Title	Advantages of YLF host over YAG in power scaling at cryogenic temperatures: direct comparison of Yb-doped systems
Subject and Keywords	info:eu-repo/classification/ddc/620
Source	Optical materials express 12(7), 2508 - 2528 (2022). doi:10.1364/OME.460445
Resource Type	info:eu-repo/semantics/article
Resource Type	info:eu-repo/semantics/publishedVersion
Publisher	OSA
Date	2022
Rights Management	info:eu-repo/semantics/openAccess
Coverage	DE
Resource Identifier	https://bib-pubdb1.desy.de/record/479017
Resource Identifier	https://bib-pubdb1.desy.de/search?p=id:%22PUBDB-2022-02875%22

INTEGRATING COST DATA INTO REPOSITORIES



DESY: Cost data workflow

PUBDB
DESY Publications Database

Home > OpenAccess
OpenAccess

Search 41,530 records for: any field

Latest additions:

2023-06-07 11:13 **OPEN ACCESS** [PUBDB-2023-03674] Journal Article
Tumasyan, A ; Adam, W ; Andrejkovic, J. W. ; *et al*
Search for nonresonant Higgs boson pair production in final state with two bottom quarks and two tau leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV
[CERN-EP-2022-117; CMS-HIG-20-010; arXiv:2206.09401]
Physics letters / B 842, 137531 (2023) [10.1016/j.physletb.2022.137531]
A search for the nonresonant production of Higgs boson pairs (ttH) via gluon-gluon and vector boson fusion processes in final states with two bottom quarks and two tau leptons is presented. The search uses data from proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV recorded with the CMS detector at the LHC, corresponding to an integrated luminosity of 138 fb^{-1} .]

OpenAccess: [PDF](#) [PDF \(PDFA\)](#) [External link](#) [Fulltext](#)

ddc: 530

Detailed record - Similar records

2023-06-07 11:13 **OPEN ACCESS** [PUBDB-2023-03673] Journal Article
Tumasyan, A ; Adam, W ; Andrejkovic, J. W. ; *et al*

OPENAPC

2 Parent Level View | Evaluation | Publisher | Journal | Filters: Year All | Parent Status All | Query All

€89,855
Optics Express

€45,476
Scientific Reports

€30,467
PLoS

€19,775
Scientific Data

€33,951
Communications Physics

€15,756
Communications Chemistry

€89,017
Nature Communications

€25,514
Communications Physics

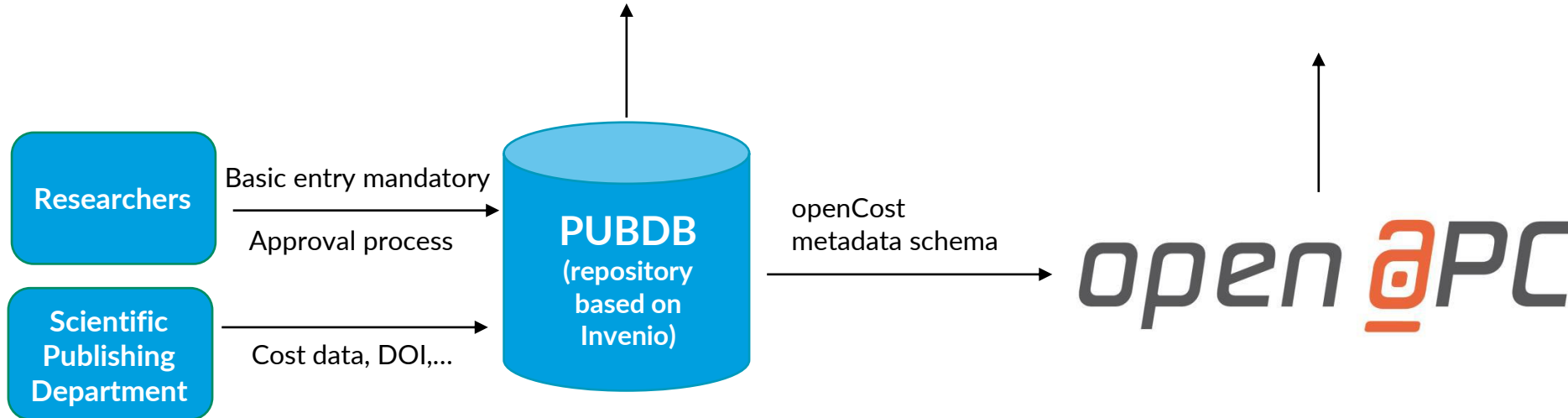
€14,264
Data

€24,814
Open Access Library

€13,910
PLoS

Measure: Sum | Number of Articles | Mean Value | Standard Deviation | Download as: CSV | JSON

DESY // Journals (87 entries)	Sum	Number of Articles	Mean Value	Standard Deviation	Percentage
Optics Express	€99,600	49	€1,834	€247	12.95%
Nature Communications	€99,017	22	€4,548	€814	12.72%
Scientific Reports	€45,476	30	€1,518	€151	7.81%
Journal of Synchrotron Radiation	€33,951	26	€1,305	€317	5.23%
Structural Dynamics	€25,514	12	€2,125	€893	3.93%

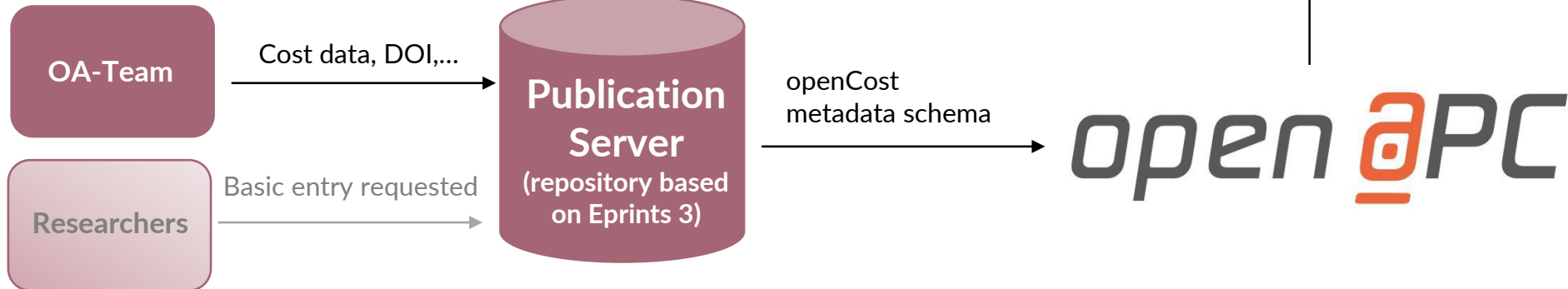


University Regensburg: Cost data workflow

University of Regensburg Publication Server

- Krickl, Patricia and Poschlod, Peter (2023) *Calcereous grassland restored by clearance and subsequent sheep grazing show fast recovery of plant functional traits - Results from a 25-year-long experiment*. *Global Ecology and Conservation* 45, e02509.
- Hense, Katharina, Deuter, Daniel, Greenlee, Mark W., Wendl, Christina, Schmidt, Nik Ole, Stroszczyński, Christian, Doenitz, Christian, Ott, Christian and Rosengarth, Katharina (2023) *Analysis of Functional Neuroplastic Changes in the Cortical Language System in Relation to Different Growth Patterns of Glioblastoma*. *Brain Sciences* 13 (6), p. 867.
- Weber, Markus, Meyer, Matthias, Von Kunow, Frederik, Flichtmeier, Bernd, Hillmann, Axel and Wulbrand, Christian (2023) *How Relevant Is the Parallax Effect on Low Centered Pelvic Radiographs in Total Hip Arthroplasty*. *Journal of Personalized Medicine* 13 (6), p. 881.
- Ettl, Tobias, Gottsauner, Maximilian, Kuhnle, Thomas, Maurer, Michael, Schuderer, Johannes G., Spörl, Steffen, Taxis, Jürgen, Reichert, Torsten E., Fiedler, Mathias and Meier, Johannes K. (2023) *The Folded Radial Forearm Flap in Lip and Nose Reconstruction—Still a Unique Choice*. *Journal of Clinical Medicine* 12 (11), p. 3636.

Measure	Sum	Number of Articles	Mean Value	Standard Deviation	Percentage
Regensburg U of Journals (298 entries)					
■ PLOS ONE	€256.842	165	€1.388	€240	13.99%
■ Frontiers in Psychology	€119.433	08	€1.716	€191	6.91%
■ International Journal of Molecular Sciences	€97.292	00	€1.622	€302	4.99%
■ Scientific Reports	€93.751	50	€1.880	€173	4.79%
■ Nature Communications	€71.441	17	€4.202	€1.151	3.91%



Bielefeld University: Cost data workflow

1757 Publikationen

1 2 3 4 ... 175 176 »

2023 | Zeitschriftenaufsatz | Veröffentlicht | PUB-ID: 2967889

Keeping the Spirit in the Bottle: On Pathological Reduction of Information in Totalitarianism
 Postoutenko K (2023)
 Information 14(1): 24.
 PUB | PDF | DOI | WoS

2023 | Zeitschriftenaufsatz | Veröffentlicht | PUB-ID: 2967891

Urbanicity: Perspectives from Neuroscience and Public Health: A Scoping Review
 Senkler B, Freymüller J, Lopez Lumbi S, Hornberg C, Schmid H-L, Hennig-Fast K, Horstmann G, McCall T (2023)
 International Journal of Environmental Research and Public Health 20(1): 688.
 PUB | PDF | DOI | Download (ext.) | WoS | PubMed | Europe PMC

Filter und Suchbegriffe

✕ Entferne Suchbegriff: funded=1

Suche

Suche Los!

Publikationen filtern

▼ Publikationstyp

▼ Publikationsjahr

▶ 1692 OA Volltext(e)

▼ Indexiert in

Darstellung / Sortierung

OPENAPC

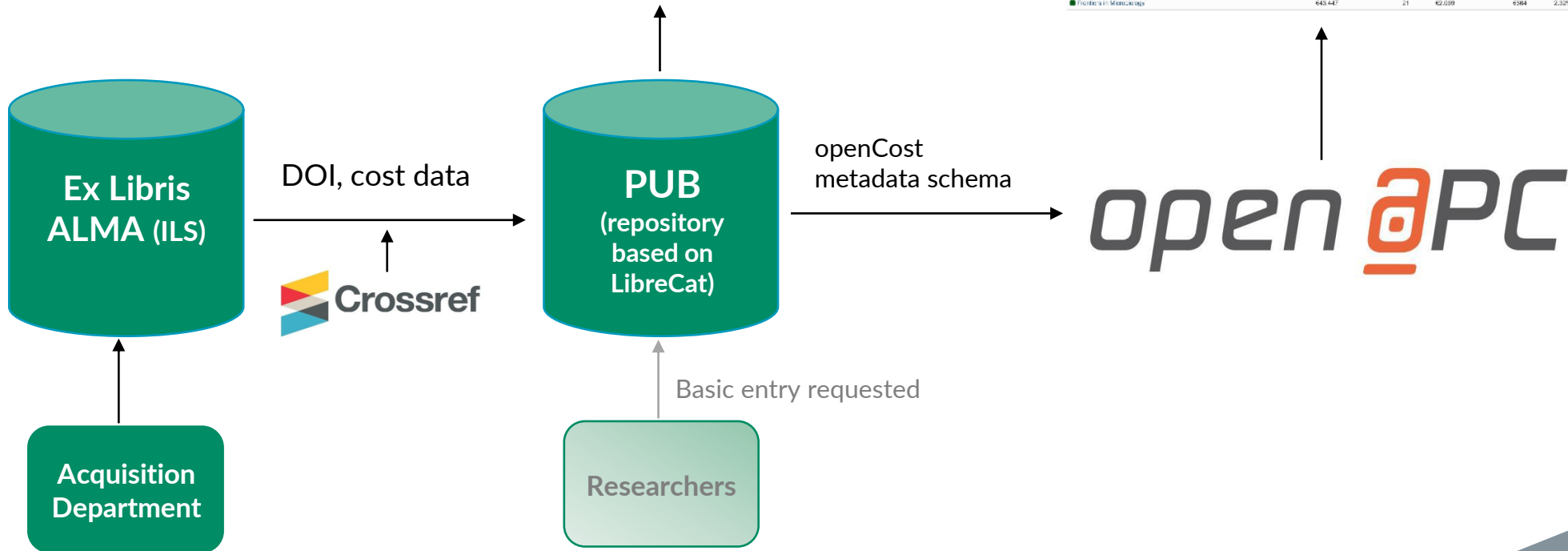
2 Publikationskategorien | 1000 | 1000 | 1000 | 1000

€169.117 PLOS ONE

€143.051 Frontiers in Psychology

€89.106 Scientific Reports

Journal	Sum	Number of Articles	Mean Value	Standard Deviation	Percentage
PLOS ONE	€169.117	110	€1.421	€385	0,09%
Frontiers in Psychology	€143.051	77	€1.858	€595	7,89%
Scientific Reports	€89.106	33	€1.851	€235	4,77%
BMC Public Health	€51.922	38	€1.366	€320	3,29%
International Journal of Environmental Research and Public Health	€58.922	27	€2.182	€390	3,59%
Frontiers in Microbiology	€43.447	21	€2.069	€384	2,32%



Invenio: Entering cost data (DESY)

[SEARCH](#) | [SUBMIT](#) | [PERSONALIZE](#) | [HELP](#) | [ADMINISTRATION](#)

PUBDB

DESY Publications Database

Home > Publications database > Highly efficient narrowband terahertz generation driven by a two-spectral-line laser in PPLN

Information
Files
Holdings

Report Total costs: 3874.79

Journal Article
PUBDB-2021-04978

Highly efficient narrowband terahertz generation driven by a two-spectral-line laser in PPLN

Olgun, H. T. (First author)^{*}; Tian, W.; Cirmi, G.^{*}; Ravi, K.; Rentschler, C.^{*}; Cankaya, H.; Pergament, M.^{*}; Hemmer, M.; Matlis, N.^{*}; Kärtner, F. (Corresponding author)^{*}

2022
Soc. Washington, DC

Optics letters 47(10), 2374 - 2377 (-) [10.1364/OL.448457]

Corresponding authors: **Olgun, Halil Tarik** *First author[*]*; **Kärtner, Franz** *Corresponding author[*]*
 Editors: ██████████@desy.de; ██████████@desy.de; ██████████@desy.de; ██████████@desy.de
 APC keys set Local Funding DFG OA Publikationskosten TIB: Optica 01/01/2023

Status	Type	Price	VAT	Currency	Cost centre	Credit card/Deposit	Date of invoice	Invoice No.	Man-ID	Comment	Local Ref.	Date
Zahlung erfolgt	Hybrid-OA	2050.15	19.14 1%	EUR	██████████		2022-04-08	448457	4484	CCBY license	52789518	2022-04-22
Zahlung erfolgt	Colour charge	597.49	5.58 1%	EUR	██████████		2022-04-08	448457	4484	first figure	52789518	2022-04-22
Zahlung erfolgt	Colour charge	1194.99	11.16 1%	EUR	██████████		2022-04-08	448457	4484	additional figures	52789518	2022-04-22
Zahlung erfolgt	Other	32.16	0.30 1%	EUR	██████████		2022-04-08	448457	4484	payment processing fee	52789518	2022-04-22
		1.00	1.00	EUR	AB-12345, Dep.		yyyy-mm-dd	123456	by pu			yyyy-mm-dd
		1.00	1.00	EUR	AB-12345, Dep.		yyyy-mm-dd	123456	by pu			yyyy-mm-dd

Save

EPrints: Entering cost data (Regensburg)

Home About Browse

Logged in as [gd@ur.de](#) | [Manage deposits](#) | [Manage records](#) | [Profile](#) | [Saved searches](#) | [Review](#) | [Admin](#) | [Edit page phrases](#) | [Logout](#)

Edit item: Article #2

Type → Upload → Details → Subjects → **OpenCost** → Deposit

< Previous Save and Return Cancel Next >

★ **Open Access Type**

- UNSPECIFIED
- APC (Gold)
- APC (Hybrid)
- VAT
- Page Charges
- Colour Charges
- BPC
- Other


Payment Details

Costs Paid:	Amount	Invoice	Type	Year	Month	Day
1.	2000	123456	UNSPECIFIED	2023	March	28
2.	380	123456	UNSPECIFIED	2023	April	24
3.			UNSPECIFIED		Unspecified	?

Costs Paid:	Amount	Invoice	Type	Year	Month	Day
1.	2000	123456	UNSPECIFIED	2023	March	28
2.	380	123456	UNSPECIFIED	2023	April	24
3.			UNSPECIFIED		Unspecified	?

< Previous Save and Return Cancel Next >

opencost is powered by EPrints 3.4 which is developed by the [School of Electronics and Computer Science](#) at the University of Southampton. [About EPrints](#) | [Accessibility](#)



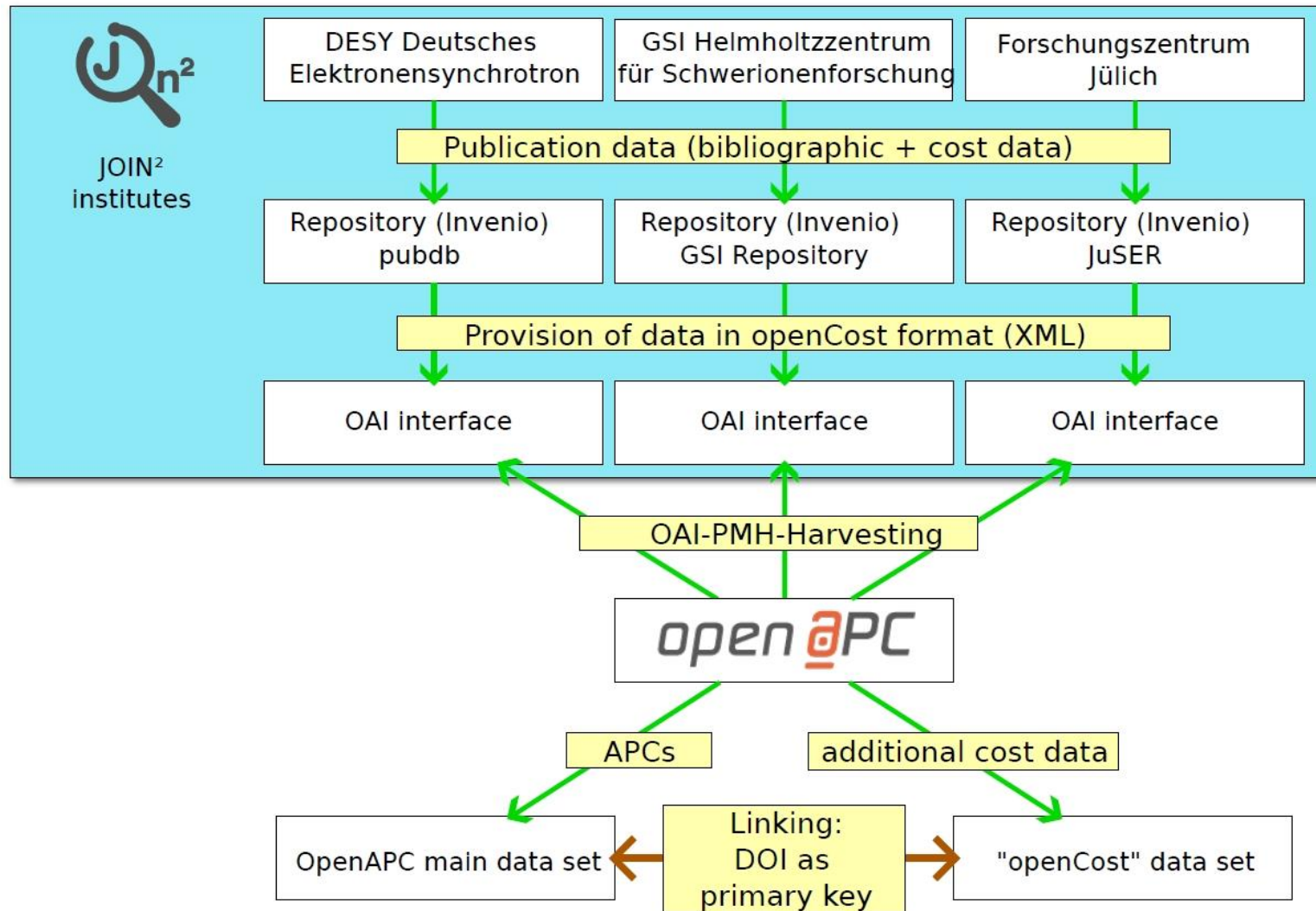
HARVESTING



OpenAPC use case: Harvesting additional cost data

- Open Data project, established at Bielefeld University Library in 2014
- Aggregates data on fee-based OA publishing
- Some institutions already integrate cost data directly into their institutional repository and letting OpenAPC harvest it via the OAI-PMH protocol
- Until now: The OpenAPC XML schema for exchanging cost data is still preliminary and not yet specified!
- Approach: By adapting to the openCost metadata schema, other cost types (apart from APCs) can also be exchanged with OpenAPC in this way and aggregated/analyzed there
 - ⇒ **Pilot Project: Integrating additional cost data into OpenAPC for three Join² institutions**

Quick overview: Harvesting additional cost data



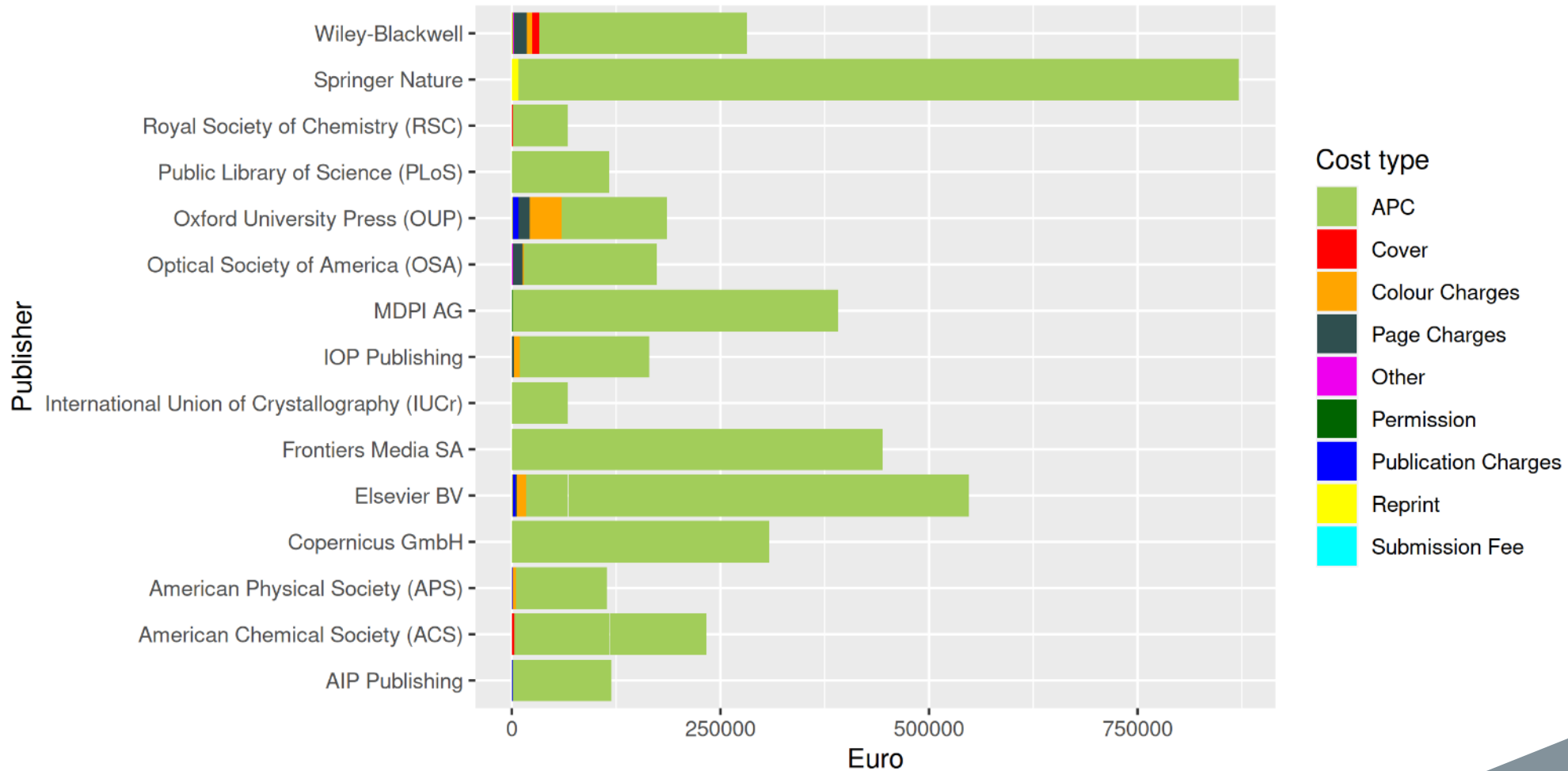
First results: Harvesting additional cost data

- A total of 2320 gold OA journal articles including cost data was provided by the 3 JOIN² institutions via harvesting
- In 268 cases, additional costs were recorded alongside an APC

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

First results: Harvesting additional cost data

- Analysis: Costs per publisher



CALL FOR PARTICIPATION

The openCost team invites you to participate in the development of the metadata schema and is looking forward to your feedback!



1. GitHub repository: <https://github.com/opencost-de/opencost>
 - Contains the schema and its documentation (including an XSD file for formal verification, and a folder with sample files that fulfill the metadata schema)
 - Tracks the evolution of the schema in a transparent way

Please feel free to open issues in our GitHub repository to provide us with your remarks and ideas!

2. Cost data survey

- Work processes and tools of national and international libraries
- Practice of publication cost monitoring
 - ⇒ Identify needs in order to take them into account in the further course of the project



<https://forms.gle/uCNV8SZDpCWcvwWU7>



Thanks for your attention!

<https://opencost.de/>

<https://github.com/opencost-de>

<https://openapc.net/>