



Hands-on Lab: Metadata schema

Expert workshop „openCost: the road to publication cost transparency”,
Oct 6, 2022, 12:30 PM

Christoph Broschinski
Julia Bartlewski
Bielefeld University Library

Working on a standardized metadata schema

- Discussing proposal for a metadata schema
 - project's proposal compared to OpenAPC schema and "Jülich" schema
 - missing elements or inaccuracies
- Defining mandatory and optional elements



OpenAPC metadata

column	description	source	required?
institution	Top-level organisation which covered the reported costs, e.g. "Bielefeld University"	none	mandatory
period	Year of APC payment (YYYY)	none	mandatory
euro	The amount that was paid in EURO. Includes VAT and additional fees	none	mandatory
doi	Digital Object Identifier	none	mandatory
is_hybrid	Determines if the article has been published in a hybrid journal (TRUE) or in fully/Gold OA journal (FALSE)	none	mandatory
publisher	Name of the publication house that has charged the fee	CrossRef	backup
journal_full_title	Full name of periodical that contains the article	CrossRef	backup
issn	International Standard Serial Number	CrossRef	backup
issn_print	International Standard Serial Number - print version	CrossRef	no
issn_electronic	International Standard Serial Number - electronic version	CrossRef	no
issn_l	Linking International Standard Serial Number	ISSN International Centre	no
license_ref	License under which the article has been published	CrossRef	no
indexed_in_crossref	indicates if the contribution is registered with the DOI agency CrossRef (TRUE/FALSE)	CrossRef	no
pmid	id for metadata records indexed in Europe Pubmed Central (Europe PMC)	Europe PMC	no
pmcid	id for articles available in Europe PubMed Central full text collection	Europe PMC	no
ut	Web of Science unique item id	Web of Science	no
url	URL to article if no DOI is available	none	backup
doaj	Indicates if the journal is indexed in the Directory of Open Access Journals (TRUE/FALSE)	DOAJ	no

Metadata schema for articles

column	description	source	required?
institution	Top-level organisation which covered the reported costs, e.g. "Bielefeld University"	none	mandatory
period	Year of BPC payment (YYYY)	none	mandatory
euro	The amount that was paid in EURO. Includes VAT and additional fees	none	mandatory
doi	Digital Object Identifier	none	mandatory
backlist_oa	Was the book published OA in the first place (FALSE) or was it already part of a publisher's backlist and became OA retroactively (TRUE)?	none	mandatory
publisher	Name of the publication house that has charged the fee	CrossRef	no
book_title	Title of the monograph	CrossRef	recommended
isbn	International Standard Book Number	CrossRef	backup
isbn_print	International Standard Book Number - print version	CrossRef	no
isbn_electronic	International Standard Book Number - electronic version	CrossRef	no
license_ref	License under which the book has been published	CrossRef	no
indexed_in_crossref	indicates if the work is registered with the DOI agency CrossRef (TRUE/FALSE)	CrossRef	no
doab	Indicates if the book is listed in the Directory of Open Access Books (TRUE/FALSE)	DOAJ	no

Metadata schema for books

„Jülich“ metadata schema, DFG "Open Access Publication Funding"

DOI	If there are several fee types per article, repeat the whole line if necessary. If no DOI is available, the data must be entered in the spreadsheet 'without DOI'.	mandatory
Name of publisher, platform, server, ...		mandatory
Publication type	Dropdown selection: journal article , book, book part, conference proceedings, dataset, preprint, software, other	mandatory
CC licence	Dropdown selection: CC BY, CC BY-SA, CC BY-ND, CC BY-NC, CC BY-NC-SA, CC BY-NC-ND, other, none	mandatory
Original currency		mandatory
Invoice amount in original currency	For articles under transformation agreements or where the DFG grant was used for membership: enter 0.00 here.	mandatory
Euro netto	Amount for Gold OA APC. For articles under transformation agreements or where the DFG grant was used for membership: enter 0.00 here.	mandatory
Tax rate		mandatory
Euro gross	Amount for Gold OA APC For articles under transformation agreements or where the DFG grant was used for membership: enter 0.00 here.	mandatory
DFG funding amount	For articles under transformation agreements or where the DFG grant was used for membership: enter 0.00 here.	mandatory
Non-eligible fee type	Dropdown selection: Colour charge, Cover charge, Hybrid OA-Fee, Page Charge, Publication Charge, Reprint, Submission fee, other	
Membership allocation	If it is not a directly assigned APC but membership related costs (e.g. scoop3), the membership is specified here. When using the " Allocation to membership" field, the "Memberships" spreadsheet is mandatory.	
Allocation to transformative agreement	Dropdown selection If the "Allocation to transformative agreement" field is filled with a value other than "none", the "Transformative agreements" spreadsheet is mandatory.	mandatory
Funding Year		mandatory
Invoice year		mandatory
Project number A55/Project ID DFG		
DFG research areas	Dropdown selection	mandatory

Spreadsheet if a DOI is available

„Jülich“ metadata schema, DFG "Open Access Publication Funding,, (2)

Additional fields, if no DOI is available:		
Title		mandatory
Author		mandatory
Title of the superordinate publication		
Volume		
Issue		
Page number		
Item number		
other identifier	e.g. URN, arxiv, PubmedID	

additional fields on the spreadsheet „without DOI“

- If records from transformative agreements or membership related costs are entered, an additional datasheet has to be filled in for each case:
- Datasheet “Membership” (11 fields: name, amount for membership, period of membership et al.)
- Datasheet “Transformative agreements” (11 fields: TA name: selection via dropdown, amount et al.)

openCost metadata schema, first proposal

```
<oc:collection>
  <oc:doi>10.1002/anie.201914888</oc:doi>
  <oc:institution>desy</oc:institution>
  <oc:rorid>01js2sh04</oc:rorid>
  <oc:invoice>
    <oc:ammount>2750.00</oc:ammount>
    <oc:fee_type>Hybrid-OA</oc:fee_type>
    <oc:currency>EUR</oc:currency>
    <oc:vat>385.00</oc:vat>
  </oc:invoice>
  <oc:invoice>
    <oc:ammount>1178.10</oc:ammount>
    <oc:fee_type>Colour charges</oc:fee_type>
    <oc:currency>EUR</oc:currency>
    <oc:vat>164.94</oc:vat>
  </oc:invoice>
  <!-- ... more cost items -->
  <oc:id_number type="local">PUBDB-2020-02650</oc:id_number>
  <oc:id_number type="pmid">32096889</oc:id_number>
  <!-- ... all known ids ... -->
  <oc:issn>1521-3773</oc:issn>
  <oc:licence>CCBY4</oc:licence>
  <oc:publisher>Wiley-VCH</oc:publisher>
  <oc:period>2020</oc:period>
  <oc:is_hybrid>TRUE</oc:is_hybrid>
  <oc:journal_full_title>Angewandte Chemie / International edition</oc:journal_full_title>
</oc:collection>
```

- XML

- Invoice element

→ Multiple entries allowed

→ fee types so far:

Gold OA-Fee, Hybrid OA-Fee, Colour charge, Cover charge, Page Charge, Publication Charge, Reprint, Submission fee, Other

→ VAT displayed separately

Comparing the different schemas

- OpenAPC schema with 5 mandatory fields vs Jülich Schema with 13 mandatory fields
- OpenAPC: CSV-files and Harvesting possible
- Jülich: Excel-files
- Open APC: 2 schemas for 2 different publication types (articles and books)
- Jülich: 1 schema for all publication types

Discussion point: costs

- openCost schema uses “invoice” elements to represent cost types – appropriate?
- What cost types do we need?
- How should we handle VAT?

Discussion point: institutions

- What are „institutions“ in the context of our metadata schema?
 - author affiliation?
 - funding institution?
 - reporting institution?
- How should institutions be designated?
 - persistent identifiers?

Discussion point: co-funding

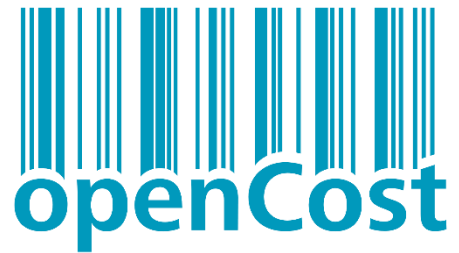
- How to deal with co-funding?
→ All-in-one record? Multiple records?
- If multiple records, how are they linked?
- How do we know if the complete costs have been reported?

Discussion point: publications types

- Which publication types/media types should be supported?
→ Surely journal articles, but there are others
- How do we want to handle media types?
→ Bibliographic metadata differs!
- Can we identify a common set of metadata fields for publications types?

Discussion point: persistent identifiers

- How do we identify publications?
→ DOI? What if no DOI exists?
- How do we identify institutions/journals/other bibliographic entities?



Thanks for your attention!

www.opencost.de