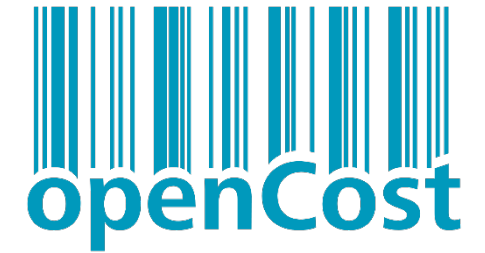




open @PC



Extension of the OpenAPC infrastructure as part of the openCost project

Julia Bartlewski, Christoph Broschinski

OpenAPC: Overview

- Open Data project, established at Bielefeld University Library in 2014
- Aggregates data on fee-based OA publishing:
 - APCs (journal articles)
 - Data on transformative agreements (since 2016)
 - BPCs (OA books, since 2020)
- Facts & numbers for the APC collection (as of 04/10/2022)
 - 381 institutions
 - 167,505 articles
 - Aggregated total APC costs: 322,494,631€

Aims & Principles

Aims

- Establish cost transparency and comparability
- Track development of costs over time
- Provide data as foundation for policy making in the ongoing OA transition

Principles

- Maximum of transparency ("Open" not only applies to data, but also to code/workflows/history)
- High degree of automation
- Low entry barriers for new participants

Data schema & workflows

- Institutions only have to provide a minimal data set
- Most bibliographic metadata is enriched by OpenAPC using external services
- Improves consistency & reduces workload for participants

institution	period	euro	doi	is_hybrid	publisher	journal_full_title	issn	issn_p	issn_e	issn_l	license_ref	in_crossref	pmid	pmcid	ut	url	doaj
Bamberg	2013	1372	10.2147...	FALSE	Dove Medical...	Journal of...	1178-7090	NA	1178-7090	1178-7090	http://creative...	TRUE	2378...	PMC3...	ut:00020...	NA	TRUE
Bamberg	2013	960	10.3389...	FALSE	Frontiers...	Frontiers in...	1664-1078	NA	1664-1078	1664-1078	NA	TRUE	2376...	PMC3...	ut:00033...	NA	TRUE
Bamberg	2013	960	10.3389...	FALSE	Frontiers...	Frontiers in...	1664-1078	NA	1664-1078	1664-1078	NA	TRUE	2384...	PMC3...	ut:00033...	NA	TRUE

Reported by institution
 Crossref
 Europe Pubmed Central
 ISSN Organisation
 Web of Science
 DOAJ

Services & Data ingestion

- OpenAPC operates the following services:
 - GitHub repository (main data storage)
 - Project blog
 - Treemap server (data visualisation)
 - OLAP server (data API)
- Participants may provide new data in different ways:
 - Via e-mail
 - By making a pull request on GitHub
 - By integrating cost data directly into their institutional repository and letting OpenAPC harvest it

Data harvesting

- Cost data is currently harvested from 3 institutions (Univ. Regensburg, Univ. Saarland, Univ. Bielefeld) via the OAI-PMH protocol
- Metadata schema not yet fully agreed upon
- Aims in context of the openCost project:
 - Switch to a new, universal openCost metadata schema for harvesting
 - Have more institutions include cost data directly into their repositories & make OpenAPC harvest it

Pilot project: Integrating additional cost data into OpenAPC

- Join² institutions have already been aggregating additional publication costs in their repositories (some of them for years)
- Other cost types besides APCs include page/colour charges, submission fees or cover charges)
- This data can already be harvested via OAI-PMH!

Pilot project: Integrating additional cost data from 3 Join² institutions into OpenAPC: Deutsches Elektronensynchrotron (DESY), Forschungszentrum Jülich (FZJ) and Helmholtzzentrum für Schwerionenforschung (GSI)

Approach

- Harvest raw data from the 3 repositories
- Filter out closed access publications and non journal articles
- Do the OpenAPC standard metadata enrichment
- Extract additional cost data from the raw records
- Create a special table for additional costs and link it to the OpenAPC main data set using the DOI as primary key

OpenAPC main data set, containing only APC costs („euro“ column)

institution	period	euro	doi	is_hybrid	publisher	journal_full_title
desy	2016	1728.73	10.1364/oe.24.021059	FALSE	Optical Society of America (OSA)	Optics Express
desy	2016	735.56	10.3390/app6090238	FALSE	MDPI AG	Applied Sciences
desy	2016	5000	10.1002/adfm.201603191	TRUE	Wiley-Blackwell	Advanced Functional Materials
desy	2016	1712.11	10.1364/oe.24.021752	FALSE	Optical Society of America (OSA)	Optics Express
desy	2016	1069.39	10.4236/jmp.2016.79088	FALSE	Scientific Research Publishing, Inc.	Journal of Modern Physics
desy	2016	1755.54	10.1364/oe.24.025169	FALSE	Optical Society of America (OSA)	Optics Express
desy	2016	1771.04	10.1364/oe.24.025582	FALSE	Optical Society of America (OSA)	Optics Express
desy	2016	912.37	10.1364/optica.3.001209	FALSE	Optical Society of America (OSA)	Optica

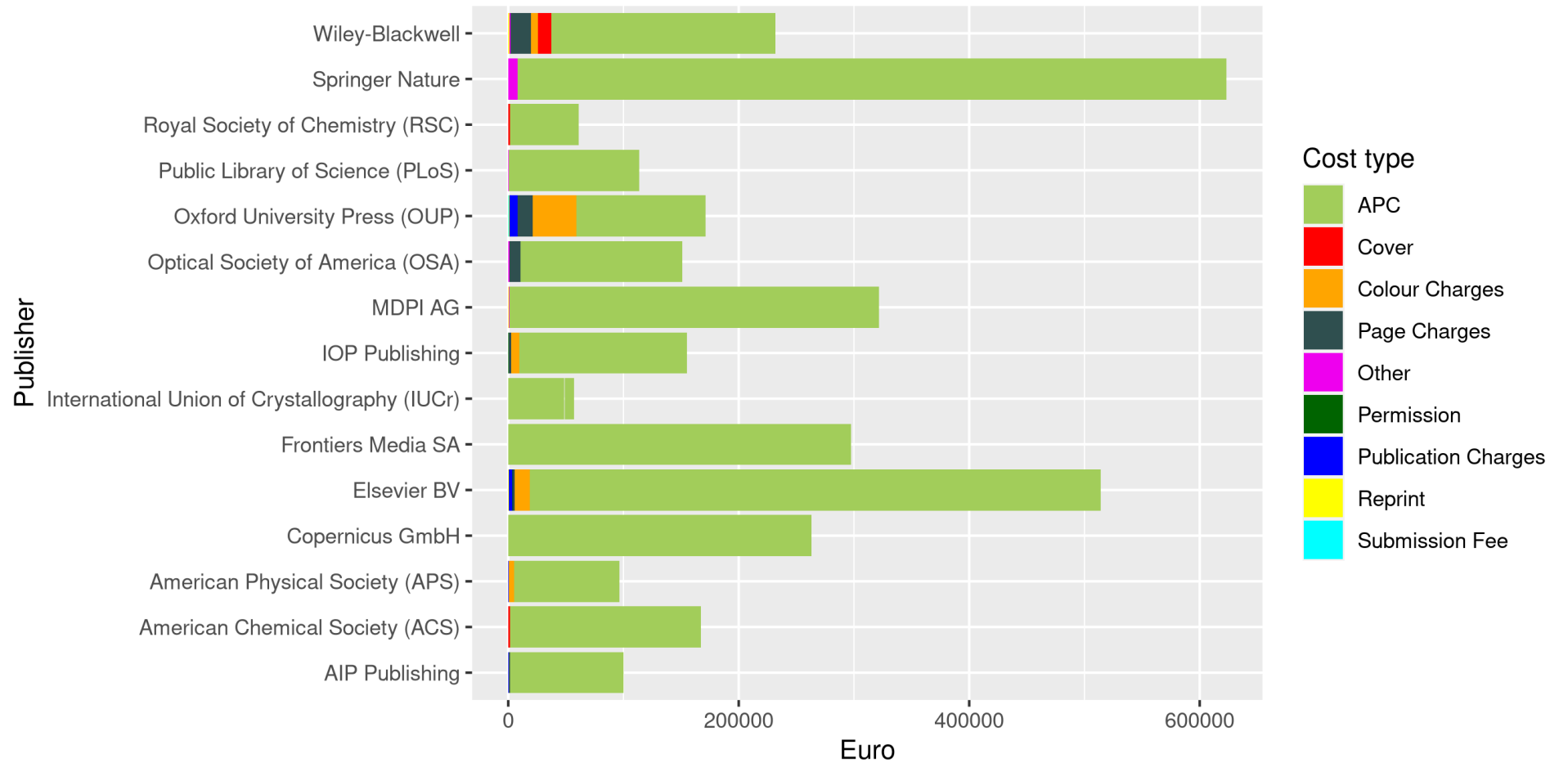
New, supporting openCost data set containing additional cost data types

doi	cover	colour_charges	page_charges	other	permission	publication_charges	reprint	submission_fee
10.1364/oe.24.021059	NA	NA	NA	NA	NA	NA	NA	NA
10.3390/app6090238	NA	NA	NA	NA	NA	NA	NA	NA
10.1002/adfm.201603191	NA	1495	NA	NA	NA	NA	NA	NA
10.1364/oe.24.021752	NA	NA	NA	NA	NA	NA	NA	NA
10.4236/jmp.2016.79088	NA	NA	NA	NA	NA	NA	NA	NA
10.1364/oe.24.025169	NA	NA	NA	NA	NA	NA	NA	NA
10.1364/oe.24.025582	NA	NA	1483.36	NA	NA	NA	NA	NA
10.1364/optica.3.001209	NA	NA	NA	NA	NA	NA	NA	NA

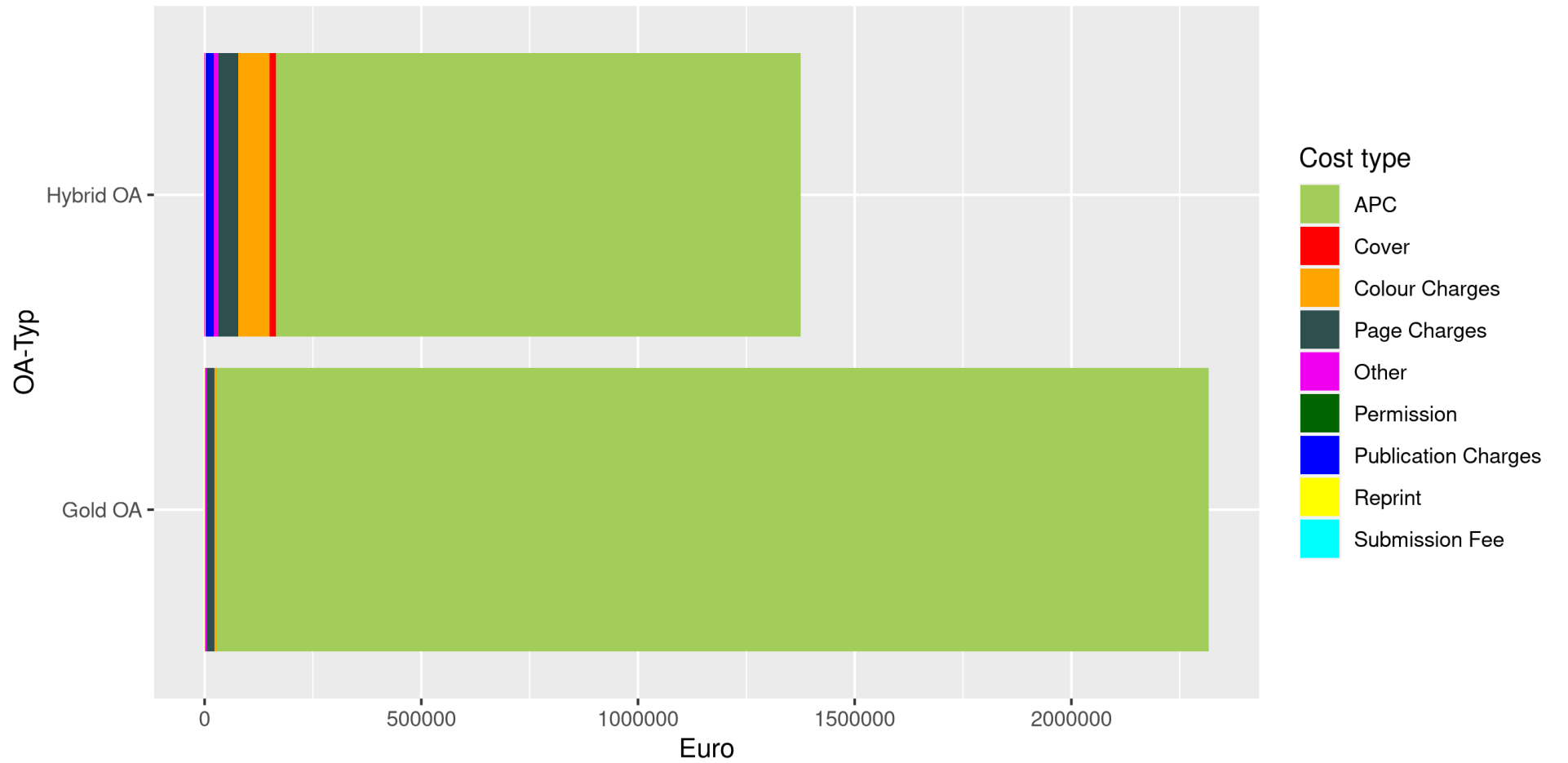
Data analysis

Cost type	Occurence	Median	Sum
APC	1933	1570	3501770
Cover	12	1206	14887
Colour Charges	41	1960	77114
Page Charges	60	838	62547
Other	100	27	13230
Permission	6	125	1098
Publication Charges	15	1143	18958
Reprint	5	399	2111
Submission Fee	10	68	768

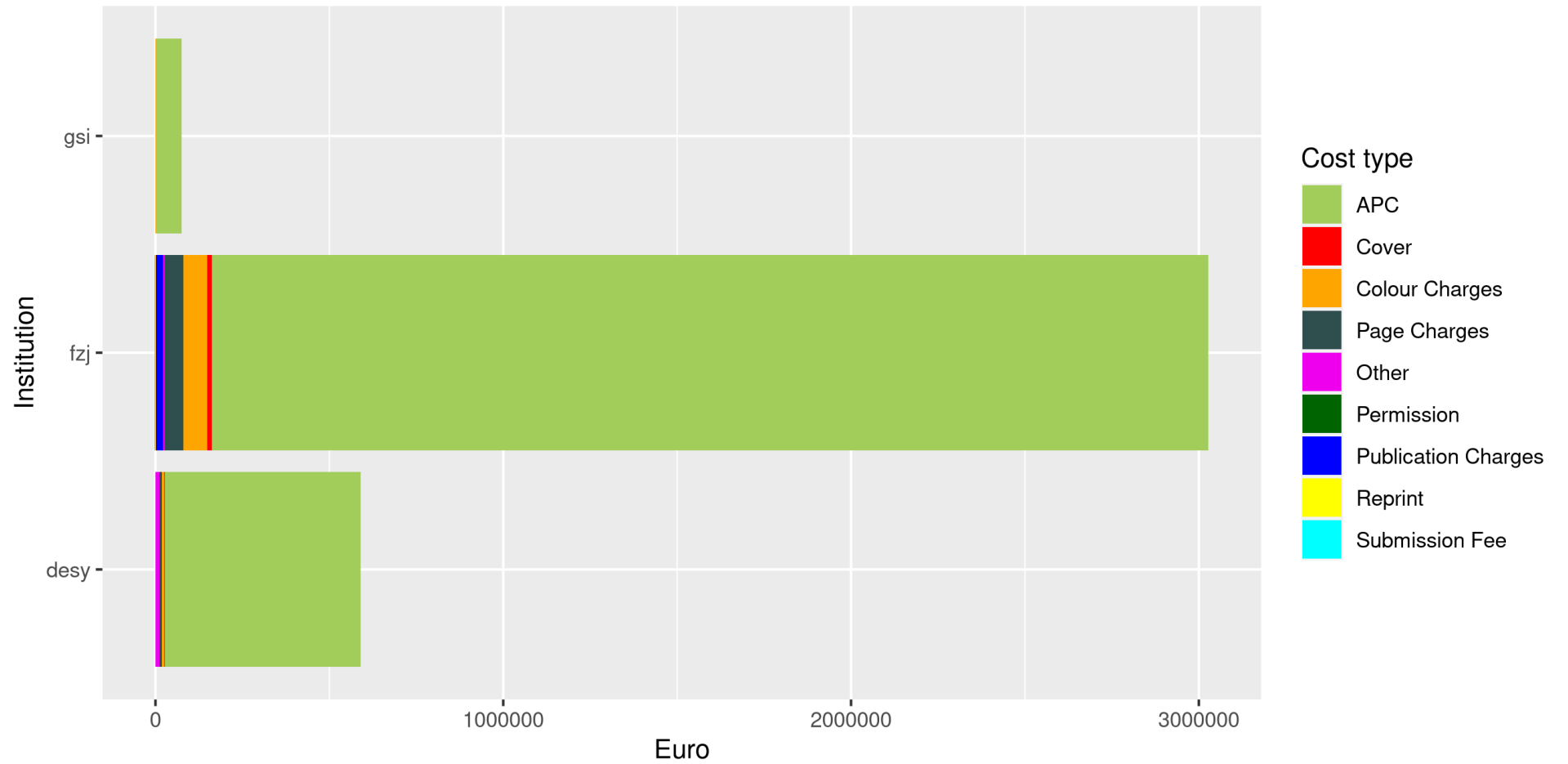
Analysis: Costs per publisher



Analysis: Costs per OA type



Analysis: Costs per institution



Results

- Aggregating additional cost data types in OpenAPC should be possible without major effort
- Integrating this kind of data could provide a more complete picture on OA publication costs
- Additional costs occur more often with certain publishers and more often with hybrid OA publishing
- But: Data is likely biased (< 2000 articles, institutions are all research institutes, publications probably STEM focused)



open @PC

Thanks for your attention!

<https://opencost.de/>

<https://openapc.net/>



UNIVERSITÄT
BIELEFELD



Universitätsbibliothek