



# OpenAPC: a contribution to a transparent and reproducible monitoring of fee-based open access publishing across institutions and nations

The OpenAPC initiative releases data sets on fees paid for open access (OA) journal articles by universities, funders and research institutions under an open database licence. OpenAPC is part of the INTACT project, which is funded by the German Research Foundation and located at Bielefeld University Library. This article provides insight into OpenAPC's technical and organizational background and shows how transparent and reproducible reporting on fee-based open access can be conducted across institutions and publishers to draw conclusions on the state of the OA transformation process. As part of the INTACT subproject, ESAC, the article also shows how OpenAPC workflows can be used to analyse offsetting deals, using the example of Springer Compact agreements.

## Keywords

Open access; OpenAPC; article processing charges (APCs); offsetting; open access transformation

DIRK PIEPER

Deputy Director  
Bielefeld University  
Library

## Introduction

In order for open access (OA) to be sustainable as the standard for academic publishing, the associated costs require monitoring. Relevant science policy strategy papers on open access and open science therefore address the issues of cost transparency and monitoring as essential success factors of a targeted OA transformation, for example the 'Amsterdam Call on Open Science'<sup>1</sup> or the 'Open Access Strategy for Germany of the Federal Ministry of Education and Research'.<sup>2</sup> In particular, the challenges lie in the creation of standardized and inter-institutional data collection and reporting routines as well as in the continuous and, as far as possible, automated quality control of the data.

CHRISTOPH  
BROSCHINSKI

Software developer  
OpenAPC

The externalization of the cost of the academic publishing system to libraries, disproportionate price increases by publishers, and complex purchasing models and confidentiality agreements have led to market inefficiency and dysfunctionality of the subscription system for academic journals.<sup>3</sup> Therefore, the transparent presentation of the costs for OA publication fees or article processing charges (APCs) is an important contribution to the reintroduction of price-limiting market mechanisms in the academic publication system. In turn, this benefits libraries, funders and authors.

'market inefficiency and dysfunctionality of the subscription system for academic journals'

The documentation of APC expenditure was significantly boosted in 2013 by the publication of corresponding data from the Austrian Fund for Scientific Research (FWF)<sup>4</sup> and in the UK

2 by the Wellcome Trust<sup>5</sup> and Jisc Collections<sup>6</sup> on the data repository figshare in 2014. Also in 2014, APC data and analyses were published on the Dataverse research platform<sup>7</sup> and Bielefeld University Library began to publish APC data on GitHub, laying the foundation of the OpenAPC project.<sup>8</sup>

These data collections have been referenced by authoritative OA transformation studies.<sup>9,10,11</sup> The various developments on the subject of OA monitoring were discussed in the context of Knowledge Exchange workshops in 2015 and 2016 and summarized in a report in early 2017.<sup>12</sup> Since 2015, OpenAPC has been funded by the German Research Foundation (DFG) within the project 'Transparent infrastructure for open access publication fees' (INTACT),<sup>13</sup> and supported by the German DINI working group 'Electronic Publishing'.

This context assured the willingness of the first German academic institutions to deliver data to OpenAPC, which motivated international funders and academic institutions to contribute to OpenAPC soon afterwards. In addition, the INTACT framework enabled synergy between the three partnering projects with the result that:

- bibliometric analyses of OA publications are conducted by the OA analytics group
- OpenAPC focuses on the aggregation of cost data on OA publishing
- the discussion about OA workflows and administrative burdens related to the management of APCs is promoted by ESAC (Efficiencies and Standards for Article Charges initiative).<sup>14,15</sup>

An early paper on OpenAPC was presented, in which self-reported spending on OA journal articles by German universities and research organizations was compared to other initiatives.<sup>16</sup> In 2018 the formerly separated data sources from FWF, Jisc and Wellcome Trust were unified within the OpenAPC data set.

## Methods, data and tools

### OpenAPC: general approach

OpenAPC follows an open science approach, which by our own standards means that everything should be visible by anyone at any time. This includes the data as well as the scripts for enrichment steps, normalization or quality checks. Using the version control system git, all files relevant to the project are kept and redacted in a repository on the platform GitHub, meaning that not only their current status, but also their complete version history is always available to the public. In addition to this, everything is automatically synchronized to a local GitLab installation at Bielefeld University Library.

'everything should be visible by anyone at any time'

### The data: structure and origins

All data accumulated by OpenAPC is willingly provided by external participants based on the principle of open data. These participants are usually called institutions in the project's parlance, although their actual nature may vary. There are data reports by individual universities or institutes, scientific organizations or research funders. Additionally, cost data may also be reported on a higher level by aggregating services (like Jisc in the UK). Taking a look at different countries, the distribution of OpenAPC institutions is described below.

#### Germany

OpenAPC started as a German national project, aiming at collecting cost data from participants of the Open Access Publishing Programme set up by the DFG.<sup>17</sup> In consequence, APC data in Germany is usually reported individually by universities (40 in total), with the vast majority of them reporting only articles funded by the said programme. German research organizations are another source of data, but again there are differences in

3 workflows: The Max Planck Society employs a central billing, where the Max Planck Digital Library (MPDL) is responsible for accounting and data reports to OpenAPC. In the case of the Helmholtz Association and the Leibniz Association, research centres operate autonomously in terms of APC payments, so they have to decide independently if they want to participate in the initiative. Altogether, 51 institutions from Germany take part in OpenAPC.

### **Austria**

In May 2016 Austria was the first country outside Germany to provide data to OpenAPC, thus extending the project's scope to an international level. Most data are reported by the Austrian Science Fund (FWF), with two participating universities completing the picture.

### **UK**

Comparable to Germany, a lot of higher education institutions in the UK are data contributors. However, those institutions are not directly in contact with OpenAPC but report their data only once to Jisc, which acts as a national aggregator<sup>18</sup> and compiles yearly collections of cost data. The Wellcome Trust represents another important source of APC data for the UK, also publishing annual reports of all their funded articles.<sup>19</sup> It is noteworthy that there is a significant overlap between the Jisc and the Wellcome data, since many institutions will also report their Trust-funded articles to Jisc. This requires a deduplication step in the OpenAPC workflows, where the Jisc data is given precedence for being more detailed on the participating institutions. In total, 51 institutions from the UK participate.

### **Sweden**

Again, there are many individual participating higher education institutions, whose data reaches OpenAPC in an aggregated form. It is noteworthy that the aggregation service is again an OpenAPC project: in May 2016 the Swedish National Library (Kungbib) launched its own survey of cost data (OpenAPC Sweden),<sup>20</sup> as at that time no comparable collection existed on a national level. The project was built in close co-operation with OpenAPC, with intensive reuse of tools and infrastructure. Data from 13 Swedish institutions are currently being incorporated into OpenAPC.

### **Norway**

In January 2018 the National Centre for Systems and Services for Research and Studies (CERES) provided the first APC data for 15 universities and research institutions in Norway for the years 2015 and 2016 in aggregated form.

### **Switzerland**

The Swiss National Science Foundation (SNSF) operates a fund to support the OA transition of all publications emanating from SNSF-funded research until 2020. The corresponding APC data was made available to OpenAPC in February 2018.

### **Italy, Spain, Canada and the USA**

There are examples of isolated participation by universities – two institutions from the US and one each from the other three countries – which often play a pioneering role within their countries with regard to open data and open access.

Altogether, as of May 2018, OpenAPC has compiled a database of 50,863 articles, with total reported costs amounting to more than 96 million euros. Figure 1 shows the evolution over time.

'OpenAPC has compiled a database of 50,863 articles, with total reported costs amounting to more than 96 million euros'

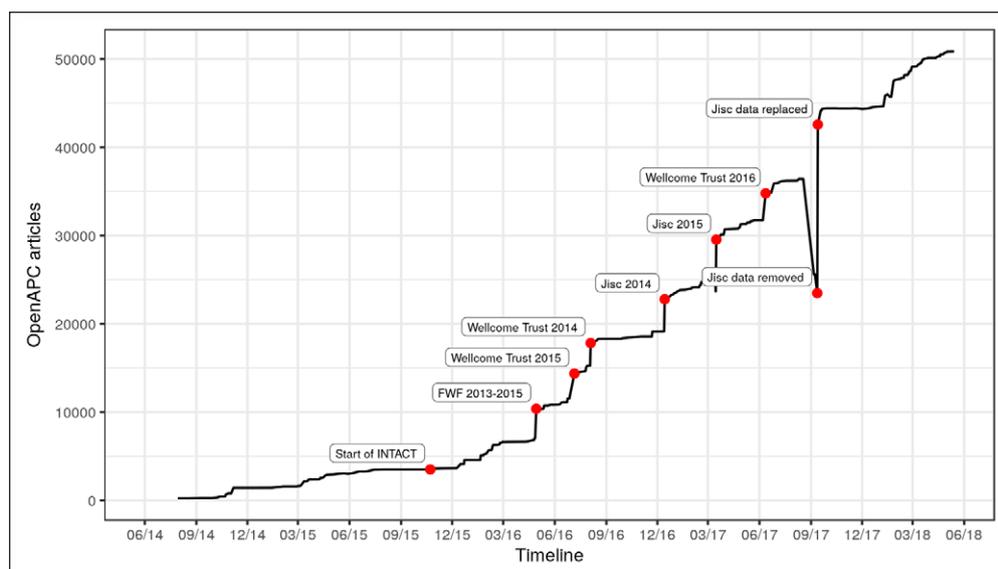


Figure 1. Development of the total number of articles in OpenAPC since 30 July 2014 (creation of the aggregated APC data file on GitHub). Major ingestion events are marked separately

## Cost data

As the project name implies, OpenAPC is intended to collect and publish data on costs incurred by institutions for publishing articles in OA journals (both hybrid and full gold OA). It is therefore very important to define what 'costs' are in the scope of OpenAPC. This attribution is less trivial than it might seem.

The first insight is that costs are not equivalent to prices. Many publishers and journals explicitly state the APCs to be paid on their web pages (so-called list prices); this information is also collected, for example, in the Directory of Open Access Journals (DOAJ).<sup>21</sup> However, experience shows that list prices usually differ from the amounts actually paid, meaning they can only be considered a rough starting point. First, it has to be taken into account that most publishers employ a dynamic price model. Institutions in the global south usually receive discounts. There are also a number of factors that may influence actual pricing. Aside from the results of individual negotiations, there may be other forms of benefit, for example due to frequency of publication, prepayment deals, society memberships or editor/reviewer activities. The latter may also lead to a publisher granting a number of 'free' articles which are published open access without any further costs.

'The first insight is that costs are not equivalent to prices'

Furthermore, APCs may need to be paid in a currency other than the institution's accounting currency, raising the question of how precisely the required conversions have been calculated. This is particularly problematic for participants from outside the eurozone who pay APCs in their domestic currency. Since no conversion takes place in those cases, exact information on the date of payment (which is necessary for precise conversion to euro amounts) is often missing. And, finally, there is even the very elementary question of whether value added tax should be included in the reported amount.

Another aspect is that there are other settlement models to pay for the OA status of a journal article. An example would be the Royal Society of Chemistry's (now discontinued) 'Gold for Gold' programme, which offered the purchase of a number of vouchers for a fixed amount, each one entitling the publication of a single OA article in a hybrid RSC journal.<sup>22</sup> On the other hand, there are offers that are in line with the APC approach but do not relate to journal articles, for example IntechOpen<sup>23</sup> publishes OA books that charge comparable fees for submitting book chapters. Confusingly, for some time the publisher explicitly referred to these fees as APCs, although this type of publication is clearly not an article according to bibliographic standards. A similar case exists with the Association for Computing Machinery, which also charges OA fees for publishing in conference proceedings.<sup>24</sup> As a final point, it should be mentioned that APCs are not the only costs

5 that may arise during OA publishing. Some journals levy additional fees for manuscript submission, while elsewhere, page and colour charges are not a thing of the past even in the age of electronic publishing.

All these questions had to be answered in order to derive guidelines for the participating institutions under the premise that cost data should be as uniform and comparable as possible but at the same time easy to collect and report. OpenAPC has developed the following policy:

- For consistency, OpenAPC only collects data on fees paid for journal articles (APCs). Other publication types such as conference papers or book chapters will not be included.
- All reported APC costs are considered 'final sums'. All modifying factors such as taxes or discounts should already be included. In other words, OpenAPC is only interested in the amount that was ultimately deducted from an institution's budget. To limit complexity, those modifiers are not included in the data set directly (see also the following section), but participants are encouraged to give more details on them as free text in an optional README file.
- The final sum principle only applies to APC costs. Additional costs such as submission fees or page/colour charges should not be included in the reported amount.
- Only articles that conform to a 'standard' APC model will be included, i.e. OA publication against direct payment. Alternative models where costs can only be calculated in hindsight (such as the aforementioned voucher system or offsetting contracts) should not be considered.
- From the previous point, it also follows that only articles with a positive APC amount should be reported. Entries with costs of zero are not included.

### Data format and enrichment

With the first delivery of cost data from an external participant (publication fund data of Regensburg University Library on 30 July 2014),<sup>25</sup> a fundamental question was the extent and scope of additional metadata to be collected. At that time only the publication of APC data from the UK by the Wellcome Trust and Jisc were available as an example. The latter was of particular interest, as Jisc acted as a national aggregator, collecting and processing cost data from external institutions. Jisc decided on a very comprehensive approach. Following the recommendations of a pilot study conducted by service provider Information Power,<sup>26</sup> the first version of the template in 2014 to be filled out by participants consisted of 34 metadata fields, with extensive bibliographic data (author, title, journal, publisher) as well as typical identifiers (DOI, PubMed ID).<sup>27</sup> However, this approach proved not to be without problems, as an analysis of the aggregated data shows. The resulting table columns are filled to varying degrees depending on the reporting institution, there are different formatting standards (dates, monetary amounts) and inconsistent designations for publisher and journal names.

As a result, the OpenAPC project employed a diametrically opposed approach: while at the beginning some bibliographic data were still required, in the end the number of mandatory data points was reduced to only five out of 18 total fields:<sup>28</sup>

- top-level organization which covered the fee (*institution*)
- year of payment (*period*)
- APC amount (*euro*)
- article DOI (*doi*)
- a Boolean indicator if the journal is hybrid or gold OA (*is\_hybrid*).

Only for those articles without a DOI, four more fields are mandatory:

- publisher (*publisher*)
- journal title (*journal\_full\_title*)
- International Standard Serial Number (*issn*)
- a link to the article full text or landing page (*url*).

6 The nine remaining fields are not required:

- ISSN for print version (*issn\_print*)
- ISSN for electronic version (*issn\_electronic*)
- linking ISSN (*issn\_l*)
- a Boolean indicator if the DOI is indexed in Crossref (*indexed\_in\_crossref*)
- the licence under which the paper has been published (*license\_ref*)
- PubMed ID (*pmid*)
- PubMed Central ID (*pmcid*)
- Web of Science unique item ID (*ut*)
- a Boolean indicator if the journal is listed in the DOAJ (*doaj*).

All non-mandatory fields of the OpenAPC data set are automatically enriched from external sources via scripts, specifically Crossref, Europe PubMed Central, DOAJ, Web of Science and the ISSN organization. The first three offer public APIs, while requests to Web of Science are restricted to members. The ISSN organization does not provide a distinct API; for every enrichment an updated mapping table has to be downloaded manually instead. Figure 2 shows all steps of the enrichment process.

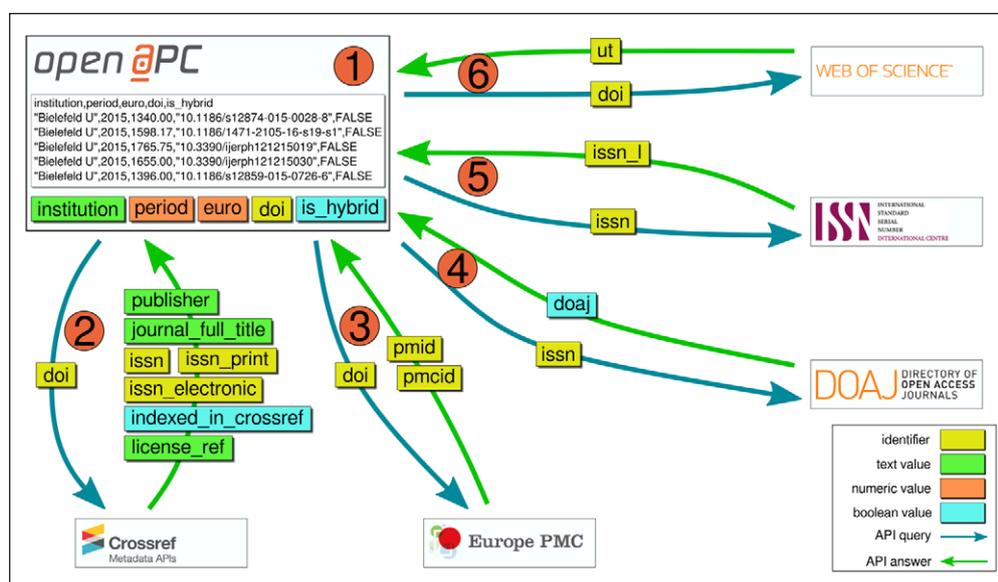


Figure 2. Overview of the OpenAPC metadata enrichment. Note how the existence of a single DOI in the input data is sufficient to bootstrap the whole process

This approach has a number of advantages:

1. The workload for data from supplying institutions remains manageable, since only the three data points – costs, DOI and journal type – have to be determined for each article. At the same time, a simple format lowers the entry threshold for new participants.
2. The automatic enrichment ensures consistent assignments of publisher names and journal titles, which is very important for later evaluations and visualizations.
3. Input data are normalized and reformatted during the enrichment process so that the results always conform to the OpenAPC data schema.
4. Corrections to secondary identifiers (ISSN-L, PubMed IDs, Web of Science identifiers) or licence information can be automatically included for the entire data set at regular intervals.

The enrichment process itself is also subject to the open data principle. Every submitted file is stored as an unmodified original in the institution's data directory on GitHub. The enriched result is then added as a second file (usually marked by the '\_enriched' suffix), making input and output comparable. The enrichment scripts are placed under an open source licence (MIT License)<sup>29</sup> and are also made public on GitHub.

Finally, the content of all enriched files is aggregated into a main CSV file (the *core data file*), which represents the OpenAPC data set.<sup>30</sup>

## 7 Automated data verification

All data reported to OpenAPC have been manually created and combined at some point in their life cycle. It is thus inevitable that the reports contain errors. Typing and copying mistakes (especially problematic in connection with DOIs), flawed formatting of monetary values or erroneous assignment of journal hybrid status are some examples. Some of these issues already get fixed during enrichment, where, for example, non-resolving DOIs are logged for review or incorrect journal titles are overwritten by imports from Crossref. This, however, is not enough. Errors at the semantic level, such as duplicate entries or inconsistencies in journal designations, cannot be resolved this way and it also cannot be guaranteed that the external metadata themselves are correct in all cases.

For this reason a small programme was written to automatically check the whole OpenAPC data set for errors.<sup>31</sup> From a formal point of view, this is a test suite as it is usually employed in software development, although the principle has been turned upside down. While predefined data are commonly used in such setups to test variable functions, here predefined functions are used to test variable data (namely, the articles in the OpenAPC data set).

The general principle is that every entry must pass a set of tests, both individually as well as interdependently (i.e. tests against each other article). The following properties are checked:

- each row has to be composed of exactly 18 columns
- publisher and journal names may not be empty or unknown (NA)
- all Boolean variables (*is\_hybrid*, *indexed\_in\_crossref*, *doaj*) must either be TRUE or FALSE
- all values in the *doi* column must represent a formally valid DOI (tested using a regular expression). If the DOI is unknown (NA), the *url* column may not be empty. No DOI may occur more than once
- the *issn* column may not be empty or NA. Its value is checked both syntactically (regular expression) and semantically (ISSN check digit calculation) if it represents a formally valid ISSN. The other ISSN columns may be empty, but if they are not, they must pass the same checks
- the value in the *euro* column must represent a numerical value larger than zero (no thousands separator; dot as decimal mark)
- if the *doaj* column is TRUE, the *is\_hybrid* column must be FALSE. (The DOAJ only lists fully open access journals)
- articles with identical values in at least one of the *issn*, *issn\_print* or *issn\_electronic* columns must also be identical in the *is\_hybrid*, *journal\_full\_title* and *publisher* columns. This test is not always reliable since title, publisher or hybrid status may change over the course of time. In those cases, ISSNs can be whitelisted to skip this set of tests.

In its primary work mode the test script can be executed on a local machine to verify any changes made to the central APC file before pushing them to GitHub. In addition, the code is also bound to a continuous integration service (in our case: Travis CI).<sup>32</sup> This web-based service monitors the OpenAPC repository, calls the test routines whenever a change occurs (a so-called build) and makes the results publicly accessible. While this may seem redundant as it just repeats the local tests, it has two distinct advantages. Firstly, it puts the open data principle into practice once more. A user can see the integrity of the OpenAPC data set at first glance (since a small widget on the main OpenAPC page displays and links to the latest test results). Secondly, it creates historical context, since test results of previous builds are also kept accessible. For an example, one may look at an early build created on 23 June 2016.<sup>33</sup> At this stage the data set contained several errors because some articles included neither a DOI nor a URL. (The corresponding rule was not in place at the beginning of OpenAPC.)

'A user can see the integrity of the OpenAPC data set at first glance'

## Dynamic documentation

In the previous sections it has been shown how automated scripts and routines support OpenAPC during data ingestion and verification. In the following sections we will shift the focus to a third aspect, which is usually more important to data reusers: dissemination and representation. The OpenAPC data are collected in a CSV file, which means that on the one hand it is highly compatible and easily processable by a wide range of tools and programmes, but on the other, not really suited to human readers. To tackle this problem, one of our first steps was the creation of a descriptive page which provides information on the current state of the OpenAPC data set. For instance, some basic statistics like the total number of articles, total sum of costs or number of participating institutions, but also advanced figures like a graphical plot showing the development of average costs over the course of time. This representation is realized as a Markdown file<sup>34</sup> and displayed on the main page of the OpenAPC GitHub repository. (If a file called README.md is present in a directory, GitHub tries to render it below the file tree per convention.) While this solution is a good way to disseminate some basic numbers about the project, it comes with its own problems. Since the OpenAPC data set is prone to changes, the information on the page will become outdated very quickly, requiring time-consuming recalculations and edits to the Markdown file. Fortunately, there is an elegant solution to this problem: the usage of dynamic reporting. This concept means that a document is not maintained as a static entity, which can only be edited by a human user, but instead it is generated from a template file, where small, interwoven chunks of programming code generate all the dynamic parts directly from underlying data. In our case, the generating template<sup>35</sup> is another Markdown file, with the code parts being written in the statistical programming language R (thus the template's *.Rmd* file ending, meaning 'R Markdown'). As it is easy to see, the template closely resembles the README file, but wherever a number, table or plot is meant to appear in the result, a code snippet can be found instead, which will produce the according entity directly from the current version of the OpenAPC data set. The generation process itself is realized by an R package called knitr.<sup>36</sup> This concept hails from the paradigm of reproducible research,<sup>37</sup> which can be seen as a subtopic of open science.

Dynamic reporting also comes into play in OpenAPC's project blog,<sup>38</sup> the main channel to disseminate information about new data contributions. Technically, the blog is another git project,<sup>39</sup> with posts being written in Markdown and then transformed into regular HTML with Jekyll (done automatically by the underlying hosting platform, GitHub pages). Since most blog posts also contain several elements which are directly dependent on OpenAPC data (both the main data file and the latest enriched file contributed by an institution) and at the same time are quite uniformly structured, it is an obvious solution to employ the same dynamic reporting techniques for them. In practice, for every new blog post an individual R Markdown file is derived from a generalized template by filling in the necessary information (institution, URLs, date, contact person, data file links) and then again knitr is used to generate a Markdown file with all numbers and plots from it. (In the project directory all R Markdown templates are stored in the *Rmd* folder, the *posts* folder holds the generated results.) This workflow makes it possible to create many standardized yet individual and informative blog posts for every data contribution in a short amount of time.

## OLAP and visualization

While the dynamic README file and the OpenAPC blog provide a general framework for dissemination, they cannot solve two remaining issues: firstly, the OpenAPC data are difficult to reuse without proper tool support and, secondly, whilst both the blog and the README page offer certain statistics and plots, they are of no use if investigations are to be conducted which go beyond their scope. For example, even a simple question like, 'What is the average APC the University of Cambridge paid for Elsevier journals in 2016?' would already require downloading the whole OpenAPC CSV file, opening it in a spreadsheet programme and applying multiple filter operations. To solve this problem, OpenAPC set up two additional services which work in close connection: An OLAP (online analytical processing) server and a graphical front end.

9 Originating from the area of business intelligence, OLAP is, in a very general sense, a technique to organize multidimensional data (usually with a financial background) in a certain structure, providing an interface to answer certain queries. As a very simplified example, a company selling an assortment of products in different countries might want to model the proceeds of its sales in a three-dimensional OLAP system, with the type of product being the first dimension, the country the second one and the year the third one. In OLAP parlance, these dimensions would then span open a three-dimensional *cube* and the system can now provide answer to queries like, 'How much revenue did product x generate last year in all countries?' by *slicing* through it and then applying an *aggregate* function to the sliced data (in this case, a *sum*). As one can see, this example query is similar to our hypothetical question about average APCs formulated above and, as it turns out, the OpenAPC data are indeed very well suited to be modelled as an OLAP cube.

Technically, the OLAP server set up<sup>40</sup> by the OpenAPC project is based on *cubes*,<sup>41</sup> a Python-based, free software OLAP implementation. The complete data set is modelled as a single cube named 'openapc' which consists of seven dimensions:

- institution
- period
- publisher
- journal\_full\_title
- doi
- is\_hybrid
- country (this dimension cannot be derived directly from the OpenAPC data set and is added during cube creation via an institution mapping table).

In addition, the cube provides four different aggregate functions:

- apc\_num\_items (simple article count)
- apc\_amount\_sum (APC sum)
- apc\_amount\_avg (average APC)
- apc\_amount\_stddev (standard deviation of APCs).

Queries to the OLAP server can be formulated by adding URL parameters to the cube base path. The most basic query would be an aggregate of the whole cube without any data slicing.<sup>42</sup>

This will apply the four aggregate functions to all articles in the data set and return the results in JSON format. The real strengths of OLAP, however, will come into play when partitioning the data in some way. One basic operation is a *drill-down*: This operation will arrange the articles into several 'buckets' along one or more dimensions. For example, if we are interested in annual APC expenditure for all institutions, we can perform a two-step drill-down, first through the 'institution' and then through the 'period' dimension.<sup>43</sup>

This will generate a much larger answer, as OLAP will now create a lot of data subsets (one for every institution/period combination) and then again apply the four aggregate functions to all of them. Consequently, the resulting JSON structure will, for example, contain information about the number of articles reported by Bielefeld University in 2015, the sum of APCs contributed by Stockholm University for 2016 and the average APC for the University of Oxford in 2017 (and all other possible combinations). Finally, we can also show how these methods can be used to answer our original example question, 'What is the average APC the University of Cambridge paid for Elsevier journals in 2016?'. There are two possible ways here. Since our query involves three different dimensions (institution, period and publisher), we could perform another three-step drill-down through them. However, this would create an even larger answer (requiring more bandwidth), and afterwards we would have to filter out the specific result we are interested in. A better approach is to make use of a *cut* operation to slice out the exact partition of articles we are looking for.<sup>44</sup>

This returns only a small JSON object, providing exactly the information we are looking for (aggregate 'apc\_amount\_avg').

10 Querying the OLAP server is a convenient way to answer questions about the OpenAPC data set without the need to download and process the raw data, and its mode of operation (URL queries, JSON return format) means that it can also serve as a data back-end for other systems. One example for such a service is operated by OpenAPC itself, the treemaps visualisation.

Arguably the most well-known service provided by the OpenAPC project, the treemaps site<sup>45</sup> uses graphical, dynamic visualizations to display APC amounts and the percentage share of an institution's or a publisher's total amount of APCs reported to OpenAPC. It offers both individual treemaps for each participating institution (easily accessible via an interactive world map), but also aggregated collections which provide a representation of the whole data set. An interesting feature of the treemaps is their interactivity: by clicking on a rectangle, one may 'delve into' the data and have a closer look at the composition of a certain partition, down to the level of individual articles (actually equivalent to the drill-down mechanism described in the previous section, as the treemaps server will just send queries to the OLAP back-end and render the results. The cut operation, on the other hand, is realized by the filter menus in the upper right corner.). This interactive behaviour is meant to provide the treemaps with utility beyond appealing graphics: it makes it possible to explore the OpenAPC data in a simple and intuitive way, discover patterns and relations and possibly identify starting points for further investigations.

There are two additional features of the treemaps worth mentioning. Firstly, the download options on the lower right make it possible to obtain a list of all articles included in the currently selected treemap partition, either in CSV or JSON format. Secondly, the 'Data & Embed' menu generates HTML code snippets, which can be used to embed the treemap into another page; this is useful if an institution wants to improve their own site with a visualization of their APC spending without much effort.

## Basic analysis and results

The following basic analysis builds upon the OpenAPC release v3.28.7-fixed<sup>46</sup> (from 14 May 2018). The first single APC data sets originate from 2005 as this is the year that MPDL APC payments begin (see Figure 3). In this release, the data from Jisc and Wellcome Trust for 2017 are not yet provided on figshare, and will be added as soon they are available.

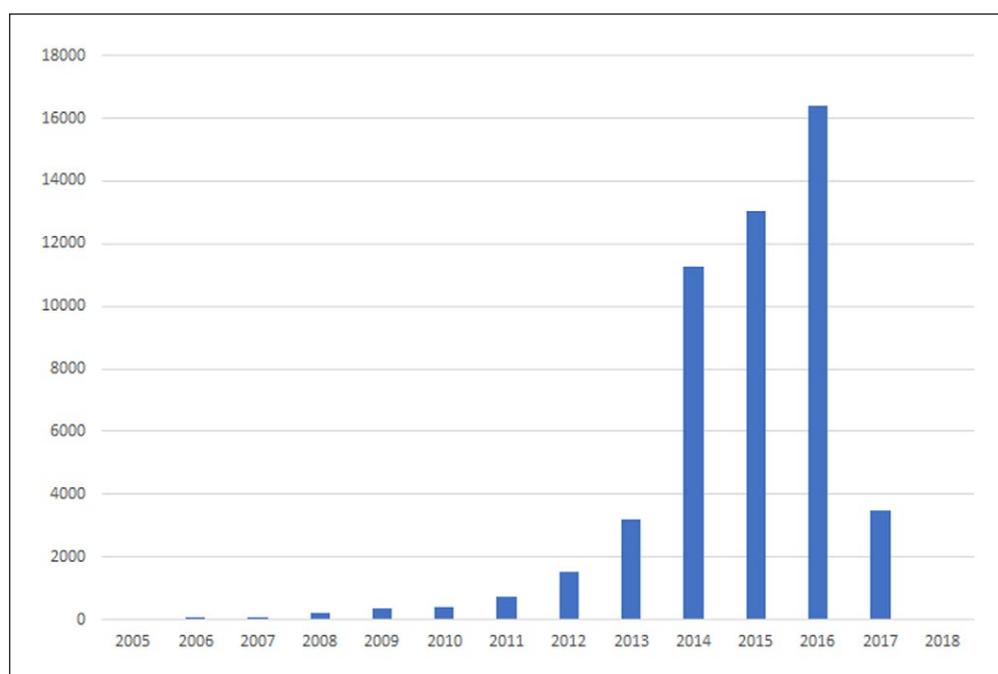


Figure 3. Number of reported APC data sets per year

Some German universities have already started to report APC data (41 data sets) for 2018. With the universities of Bielefeld and Regensburg, there are two institutions providing their data via harvesting routines of their repositories, which allows the automatic update of their

11 APC data sets in shorter intervals.<sup>47</sup> We recommend this method of data delivery in future because it decreases the administrative expenses for institutions once the OAI interface of the institutional repository is expanded while at the same time ensuring that the OpenAPC data set is up to date.

The current OpenAPC release contains 21,145 data sets for OA articles in hybrid journals with a median of €2,443 (standard deviation €929). At the same time it contains 29,718 data sets for articles in pure OA journals with a median of €1,479 (standard deviation €695). The following box-plots indicate and compare the development of the spending distribution over fully and hybrid OA journals.

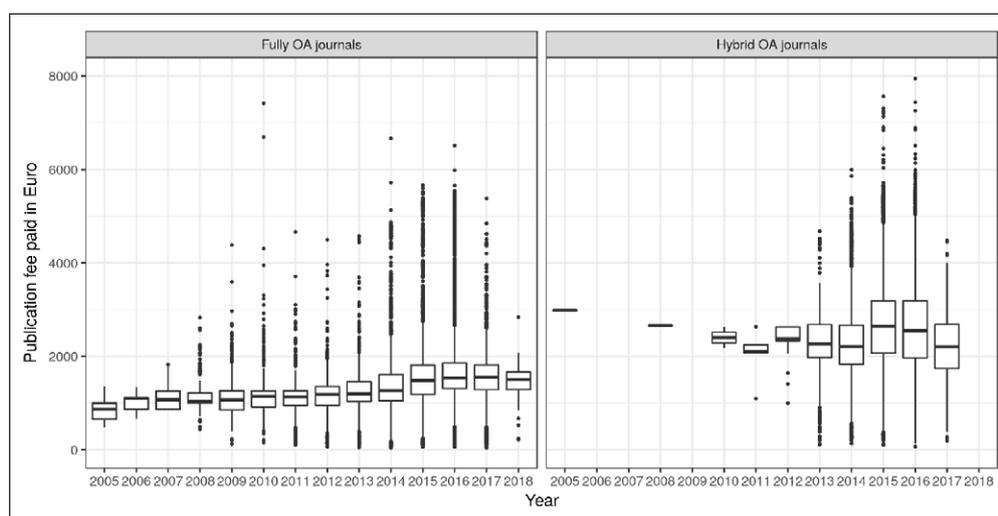


Figure 4. Spending distribution over fully and hybrid OA journals across the complete OpenAPC database

The data show very clearly the different cost levels and increases for publishing in pure OA and hybrid journals (see Figure 4). Please note that the decrease of the median for hybrid journals from 2016 to 2017 is due to the fact that the data for 2017 are not yet complete.

The OpenAPC treemap visualization allows the data set to be filtered by different dimensions, such as publishers or journals. Although the data within OpenAPC are only a sample of the publisher’s total numbers of OA articles and APC revenues from universities, research organizations and individual researchers, the sample is big enough to show what the financial flows to publishers look like. On the one hand it is no surprise that the biggest three publishers also receive most of the money spent on OA publishing, especially in hybrid journals; on the other hand we see three pure OA publishers in the top ten list, PLOS, Frontiers and Copernicus. The position of Springer Nature is a result of the merger between Springer and BioMed Central (BMC) in 2008 and Springer and Nature in 2015. (See Table 1.)

‘The OpenAPC treemap visualization allows the data set to be filtered by different dimensions, such as publishers or journals’

Publisher	Expenditure	Number of articles
Springer Nature	€19,600,685	10,964
Elsevier BV	€18,267,632	6,946
Wiley Blackwell	€10,173,721	4,353
Public Library of Science (PLOS)	€8,517,781	6,107
Frontiers Media SA	€4,626,196	3,188
Oxford University Press	€4,260,076	1,736
American Chemical Society (ACS)	€3,467,146	1,311
Copernicus GmbH	€2,300,447	1,736
British Medical Journal (BMJ)	€2,277,943	1,024
IOP Publishing	€2,232,080	1,497

Table 1. Top ten publishers according to the amount of expenditure in OpenAPC

12 The total number of OA articles in the OpenAPC data set is distributed across 1,536 pure OA and 3,329 hybrid journals. This reflects the fact that there are more subscription than pure OA journals in the marketplace. Although OpenAPC indicates much more money is spent on APCs in hybrid journals (€52,598,824) compared with that spent for APCs in pure OA journals (€43,995,077), the top ten list of journals according to the amount of APC expenditure within the OpenAPC data set now shows only pure OA journals. (See Table 2.) Although the DFG has a policy to fund articles in pure OA journals only, this is a surprising result to us, as the cost data for only 16,664 out of 50,863 OA articles in total were delivered by German universities and research organizations (32.8%). Of the 16,664 articles from Germany, 224 were hybrid (1.34%), which shows the effectiveness of the DFG policy.

Journal	Expenditures	Number of articles	oa/hybrid
<i>PLOS ONE</i>	€6,547,118	5,156	oa
<i>Scientific Reports</i>	€2,354,821	1,695	oa
<i>New Journal Of Physics</i>	€1,227,640	1,062	oa
<i>Frontiers In Psychology</i>	€1,147,759	781	oa
<i>Nature Communications</i>	€2,932,400	693	oa
<i>BMJ Open</i>	€779,896	453	oa
<i>Atmospheric Chemistry and Physics</i>	€673,856	417	oa
<i>Optics Express</i>	€663,872	386	oa
<i>Nucleic Acids Research</i>	€632,292	346	oa
<i>Cell Reports</i>	€707,710	170	oa

Table 2. Top ten journals according to the amount of expenditure in OpenAPC

## Other Applications: offsetting contracts

OpenAPC workflows and tools can also be applied to other areas of fee-based OA publishing. This section describes a side project which was set up in close co-operation with the ESAC initiative, a collection of articles published under offsetting contracts.

### Offsetting

In addition to the publishing of articles in pure OA or hybrid journals, some library consortia started to finance OA articles in hybrid journals through so-called offsetting contracts. According to 'The Joint Understanding of Offsetting', which was formulated in 2016, the offsetting models are regarded 'as transitional models in order to pave the ways to a fully open access business model'.<sup>48</sup> There are two types of offsetting agreements:

- 1) Pure offsetting agreements simply offset the subscription expenditures to a certain publisher against the APCs, which are paid to publish open access in the related journal portfolio. The motivation behind it is to mitigate the infamous double dipping, where an institution pays APCs to publish in a hybrid journal but still has to subscribe to the journal nonetheless, since it is topically relevant to their researchers as a whole.
- 2) In 'read & publish' models, the licensees pay fees for reading and for publishing open access in a defined set of journals with the intention of reducing the reading fees over time, in order to finance more and more OA articles in those journals.

The goal of those transitional agreements is to establish a pay-as-you-publish model, in which no more subscription costs or reading fees are paid and every article is open access. Because of diverging preconditions and cost differences, and due to the overall complexity of those agreements, it is difficult to determine the amount of money which might have been transformed from the subscription side of Springer Nature to the OA publishing side, especially in those cases in which the overall costs for the agreement increased significantly.<sup>49</sup> But analysing the bibliographic data can give some indication of whether the other goal of OA transformation, to flip journals into open access, can be reached via offsetting.

## Offsetting data set

As a result of the first ESAC Offsetting Workshop in 2016, the collection of articles published under offsetting contracts has been established as a side project of OpenAPC. Data providers include the Austrian Academic Library Consortium (KEMÖ), the MPDL, the Association of Universities in the Netherlands (VSNU/UKB) for all Dutch universities, the Bibsam Consortium for Sweden and Jisc Collections for the UK.

Most of the articles originate from the Springer Compact agreements, and the consortia usually report their data to OpenAPC in regular intervals. The data is then enriched and checked for errors just like regular contributions, compiled into a separate core data file (*offsetting.csv*) and published in a dedicated data subdirectory on GitHub.<sup>50</sup> The data format is the same as in the main APC file, with one notable exception. Since the offsetting articles do not have a directly attached cost (the final costs on a per-article base of such contracts can usually be calculated in hindsight only), the *euro* column is always empty. This requires some adaptations when it comes to data dissemination: the offsetting data is provided in a separate OLAP cube, which uses a reduced aggregation model.<sup>51</sup> Since there is no monetary information, the calculation of sums or averages does not make sense, so a simple article count is used instead. The same is true for the associated treemap representation.<sup>52</sup> Without cost data, the number of articles has to be used as metric for the rectangle generation.

'analysing the bibliographic data can give some indication of whether the other goal of OA transformation'

### Analysis: Offsetting coverage

While many calculation metrics developed for the standard OpenAPC data set cannot be applied to the offsetting data due to missing cost information, it is still possible to gather interesting insights from the data set. In order to understand how the articles financed through these offsetting contracts have increased the OA shares in Springer Compact journals and how offsetting is contributing to the goal of the OA2020 initiative to flip journals from the subscription system into open access, we tried to answer three questions for every journal appearing in the offsetting data set:

- How many articles have been published in this journal in a distinct period?
- How many of those were published open access?
- How many of those were published under offsetting contracts?

There was a distinct reason why looking into these questions could be a rewarding endeavour. Since we had all existing Springer Compact contractors contributing to the data set,<sup>53</sup> it could be assumed to be almost complete, so the results should provide significant insight.

To tackle these questions, it was necessary to link the offsetting data to an external source, since the publication numbers and OA shares of a journal represent bibliometric information not contained in our data. Two possible services were available: Crossref and SpringerLink, the publisher's own web portal. Both turned out to be not ideally suited for this task. Crossref has potential completeness issues and the OA status of articles has to be derived indirectly from the appended licence information, which means that there is a heavy dependency on publishers to annotate their articles correctly. SpringerLink, on the other hand, is not designed to be machine-readable, so the required journal metrics like yearly total article numbers and OA articles are not exposed directly via an API.

The most challenging issue, however, turned out to be the matching of publication years: the 'period' field in the offsetting data set relates to the acceptance date of the article as provided by Springer. Unfortunately, this type of date information is used nowhere else. It is not reported on Crossref for Springer articles, and the publisher itself refers to another time frame for the publications on SpringerLink, namely the print publication dates. In the end it was necessary to convert the whole offsetting data to a new temporal reference system by

14 looking up every single article on SpringerLink and importing the according print publication year from there. While this operation was time-consuming (even with programming support) and had the disadvantage that the resulting data set is not comparable to the original offsetting data in the period field any longer, the results justified the effort: a dedicated OLAP cube holds the results,<sup>54</sup> and an associated treemap<sup>55</sup> tabulates them. More technical details and preliminary analyses have been reported in the OpenAPC project blog.<sup>56</sup>

## Results

Compared to our first analysis of the Springer Compact agreements in our blog in March 2018, the data set has now been updated with data from KEMÖ in April 2018. Together with the other data providers, MPDL, VSNU/UKB for all Dutch universities, the Bibsam Consortium for Sweden and Jisc Collections for the UK, the offsetting collection for Springer Compact journals now contains 19,296 OA articles from 2015 to 2018. The following updated analysis also builds on the version v3.28.7-fixed and is limited to the years 2016 and 2017, as the data for the above-mentioned licensees are most completely available for these two years. In total, 14,110 articles were placed in open access through the offsetting contracts and the period mentioned above. This corresponds to 4.45% of the total number of articles in the Springer Compact journals (317,318) during this period. We were able to find a total of 26,713 OA articles in the Springer Compact journals, which corresponds to a share of 8.42%. Thus, offsetting was responsible for a little under half of all OA articles in hybrid Springer Compact journals. There was no single journal title which reached an overall OA article share above 50% for 2016 and 2017 together. The following Table (Table 3) shows all hybrid Springer Compact journals with OA shares greater than 50% in 2016 or 2017.

Journal	Year	Total number of articles	Number of open access articles	Number of offsetting articles	Open access share
<i>Psychotherapie Forum</i>	2017	20	19	1	95.00%
<i>Gynecological Surgery</i>	2017	30	28	2	93.33%
<i>Integrating Materials</i>	2017	23	19	1	82.61%
<i>Innovative Infrastructure Solutions</i>	2017	68	54	1	79.41%
<i>Cambridge Journal of Evidence-Based Policing</i>	2017	12	8	1	66.67%
<i>Journal of Remanufacturing</i>	2017	13	8	1	66.67%
<i>Journal of World Prehistory</i>	2017	13	8	7	61.54%
<i>medizinische genetik</i>	2017	44	26	1	59.09%
<i>Artificial Intelligence and Law</i>	2017	24	14	11	58.33%
<i>Liverpool Law Review</i>	2016	12	7	6	58.33%
<i>Wiener klinische Wochenschrift Education</i>	2017	9	5	5	55.56%
<i>The European Physical Journal H</i>	2017	22	12	5	54.55%
<i>Ambio</i>	2017	127	67	10	52.76%

Table 3. Springer Compact journals with an OA share of more than 50% in 2016 or 2017

In 2017 at least one offsetting article was published in a total of 1,347 Springer Compact journals, of which only 13 journals achieved an OA share of greater than 50%. Only 0.76% of the Springer Compact journals achieved an OA share of greater than 50%. If one assumes 1,700 Springer Compact journals altogether, no offsetting articles were published in around 350 titles in 2017. In another 281 journal titles, we recorded only one offsetting article. Furthermore, we observe strong fluctuations within individual journals, especially if the number of published articles per journal is low overall. In 2017, for example, the journal *Psychotherapy Forum* had an OA share of 95%. Of these, only 5.26% were a result of offsetting. In 2016 the same journal had an OA share of 10.71%, of which 100% was financed by offsetting. The OA share in journals with more than 1,000 articles per year averages about 3%, clearly below the overall OA share of all Springer Compact journals.

- 15 The first interim conclusion is that offsetting has contributed to a significant increase in OA articles in some of the hybrid Springer Compact journals from 2016 to 2017. So far, the numbers and distribution of additional OA articles generated through the above-mentioned offsetting agreements are not yet sufficient to flip individual journals completely into open access.

## Discussion

OpenAPC clearly demonstrates that a transparent and reproducible monitoring on fee-based OA publishing across institutions and nations is possible. Because the institutional or national aggregation of APCs cannot reach completeness, the aggregation and normalization of APC data within OpenAPC creates a comprehensive and statistically valid general data set on APC expenditure. It allows the analysis of cash flows from universities, other research organizations and funders to publishers and journals over time. In order to analyse cost increases over time, we recommend starting with the year 2015. After 2015, data seems more complete and therefore large enough. In the case of offsetting, it provides useful information for future negotiations in relation to article distribution, the appropriateness of price levels for certain journals, and on the question of how to reach the overall goal of OA transformation.

'offsetting has contributed to a significant increase in OA articles in some of the hybrid Springer Compact journals'

OpenAPC will not and cannot replace national or institutional reporting requirements on APC expenditure, however. Different OA strategies and funding policies in various countries have to be kept in mind when comparing the OpenAPC data by country. In the case of Germany, for example, the reported data are heavily influenced by the DFG policy to fund articles in pure OA journals only, with an additional price cap of €2,000. In the UK, about 75% of the reported APC expenditure is going on hybrid journals as a result of the implementation of the Finch report and resulting funder policies. Because the amount of data is not yet sufficiently widespread for many countries with strong publication outputs, country analysis is limited to some extent. Furthermore, OpenAPC has collected only data from well-funded countries so far; data from countries in the global south are missing. The INTACT report 'Publications in gold-open-access-journals on the global and European level and in research organizations' clearly indicates that there is APC expenditure in those countries as well.<sup>57</sup>

'it provides useful information for future negotiations'

With regard to offsetting, it seems that big deals are increasing the level of open access in hybrid journals, but not enough to flip journals entirely. Although the number of existing Springer Compact agreements is relatively small, the agreements do allow the members of 294 universities and other research institutions in Austria, Germany (N.B. the MPG agreement only allows the members of 83 MPG institutes to publish open access in Springer Compact), Netherlands and UK to publish open access in Springer Compact journals. The offset analysis also shows that there is already additional money for hybrid OA publishing in Springer Compact, which is paid by individual researchers. The still low overall OA share for Springer Compact journals might also be interpreted as a low demand for OA publishing in this journal portfolio. We are therefore looking forward to the DEAL negotiations in Germany, which may allow the members of all German universities to publish open access in Springer Compact journals, which will significantly increase the size of the data set.

### Acknowledgements

The authors would like to thank Najko Jahn and all members of the INTACT project (<https://www.intact-project.org/contact/>), the members of the DINI working group Electronic Publishing, and all data contributors (<https://github.com/OpenAPC/openapc-de#contributors>) to the OpenAPC initiative.

### Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other *Insights* articles can be accessed here – click on the URL below and then select the 'full list of industry A&As' link: <http://www.uksg.org/publications#aa>.

### Competing interests

The authors have declared no competing interests.

### Funding

The authors gratefully acknowledge the financial support of the German Research Foundation (DFG) for this project and we are looking forward to the funding of the Federal Ministry of Education and Research (BMBF) beginning in October 2018.

## References

1. Amsterdam Call for Action on Open Science, 2016:  
<https://www.government.nl/binaries/government/documents/reports/2016/04/04/amsterdam-call-for-action-on-open-science/amsterdam-call-for-action-on-open-science.pdf> (accessed 29 August 2018).
2. Open Access in Deutschland – Die Strategie des Bundesministeriums für Bildung und Forschung, 2016, BMBF:  
[https://www.bmbf.de/pub/Open\\_Access\\_in\\_Deutschland.pdf](https://www.bmbf.de/pub/Open_Access_in_Deutschland.pdf) (accessed 29 August 2018).
3. McKenzie L, 'Big Deal' Cancellations Gain Momentum: An increasing number of universities are ending, or threatening to end, bundled journal subscriptions with major publishers, 2018:  
<https://www.insidehighered.com/news/2018/05/08/more-institutions-consider-ending-their-big-deals-publishers> (accessed 29 August 2018).
4. Reckling F and Kenzian M, Austrian Science Fund (FWF) Publication Cost Data 2013, 2015; DOI:  
<https://doi.org/10.6084/m9.figshare.988754.v4> (accessed 29 August 2018).
5. Kiley R, (2014b). Wellcome Trust APV spend 2012–13: data file; DOI:  
<https://doi.org/10.6084/m9.figshare.963054.v1> (accessed 29 August 2018).
6. Lawson S, Jisc Collections aggregated APC data 2014, 2014; DOI:  
<https://doi.org/10.6084/m9.figshare.1060243.v4> (accessed 29 August 2018).
7. Morrison H, Dramatic Growth of Open Access Dataverse, 2018:  
<https://hdl.handle.net/10864/10660> (accessed 29 August 2018).
8. Bielefeld University Library APC data inside the OpenAPC GitHub repository:  
<https://github.com/OpenAPC/openapc-de/tree/master/data/unibi> (accessed 12 September 2018).
9. Schimmer R, Geschuhn K and Vogler A, Disrupting the subscription journals' business model for the necessary large-scale transformation to open access, *MGPPuRe*, 2015:  
<http://hdl.handle.net/11858/00-001M-0000-0026-C274-7> (accessed 3 September 2018).
10. *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*, Mellon Foundation, University of California Libraries, 2016:  
<http://icis.ucdavis.edu/wp-content/uploads/2015/07/UC-Pay-It-Forward-Project-Final-Report.pdf> (accessed 3 September 2018).
11. Johnson R, Fosci M, Chiarelli A, Pinfield S and Jubb M, *Towards a Competitive and Sustainable OA Market in Europe – A Study of the Open Access Market and Policy Environment*, 2017; DOI:  
<https://doi.org/10.5281/zenodo.401029> (accessed 3 September 2018).
12. Stern N, Knowledge Exchange consensus on monitoring Open Access publications and cost data, *Technical report*, Knowledge Exchange, 2017:  
[http://repository.jisc.ac.uk/6648/1/KER0001\\_KE\\_workshop\\_Monitor\\_OA\\_Publications\\_and\\_cost\\_data\\_MAR17\\_v4\\_\(2\).pdf](http://repository.jisc.ac.uk/6648/1/KER0001_KE_workshop_Monitor_OA_Publications_and_cost_data_MAR17_v4_(2).pdf) (accessed 3 September 2018).
13. INTACT project:  
<https://intact-project.org> (accessed 12 September 2018).
14. ESAC initiative:  
<http://esac-initiative.org> (accessed 12 September 2018).
15. Geschuhn K and Stone G, It's the workflows, stupid! What is required to make 'offsetting' work for the open access transition, *Insights*, 2017, 30(3), 103–114; DOI:  
<https://doi.org/10.1629/uksg.391> (accessed 3 September 2018).
16. Jahn N and Tullney M, A study of institutional spending on open access publication fees in Germany, *PeerJ*, 2016, 4: e2323; DOI:  
<https://doi.org/10.7717/peerj.2323> (accessed 3 September 2018).
17. DFG – Information on the Open-access Publishing Programme, 2015:  
[http://www.dfg.de/en/research\\_funding/programmes/infrastructure/lis/funding\\_opportunities/open\\_access/](http://www.dfg.de/en/research_funding/programmes/infrastructure/lis/funding_opportunities/open_access/) (accessed 12 September 2018).
18. Lawson S, ref. 6.
19. Kiley R, ref. 5.
20. GitHub repository of the Swedish OpenAPC project:  
<https://github.com/Kungbib/openapc-se> (accessed 12 September 2018).
21. Directory of Open Access Journals:  
<https://doaj.org> (accessed 12 September 2018).
22. Gold for Gold scheme to end 2017, RSC, 2017:  
<http://www.rsc.org/journals-books-databases/librarians-information/librarians-notes/all-articles/2016/jun/g4g-scheme-ends-2017/> (accessed 3 September 2018).
23. IntechOpen:  
<https://www.intechopen.com/> (accessed 12 September 2018).
24. ACM Information for Authors:  
<https://authors.acm.org/main.html> (accessed 12 September 2018).
25. GitHub link showing the state of the OpenAPC repository on 30 July 2014:  
<https://github.com/OpenAPC/openapc-de/tree/d304d96367> (accessed 12 September 2018).
26. Woodward H M and Henderson H L, Report for Jisc Collections on Total Cost of Ownership Project: Data Capture and Process, *Technical report*, Information Power Ltd, 2014:  
<https://www.jisc-collections.ac.uk/Global/News%20files%20and%20docs/IPL-Jisc-Total-Cost-of-Ownership-Data-Capture-Report.pdf> (accessed 3 September 2018).
27. Lawson S, APC data for 25 UK higher education institutions, *Data set*; DOI:  
<https://doi.org/10.6084/m9.figshare.1305596.v5> (accessed 12 September 2018).

28. Broschinski C, OpenAPC Data Submission Handout, GitHub, 2016:  
<https://github.com/OpenAPC/openapc-de/wiki/Data-Submission-Handout> (accessed 12 September 2018).
29. Open Source Initiative – The MIT License:  
<https://opensource.org/licenses/MIT> (accessed 12 September 2018).
30. Core data file in the OpenAPC GitHub repository:  
[https://github.com/OpenAPC/openapc-de/blob/master/data/apc\\_de.csv](https://github.com/OpenAPC/openapc-de/blob/master/data/apc_de.csv) (accessed 13 September 2018).
31. Source code of the OpenAPC data test script:  
[https://github.com/OpenAPC/openapc-de/blob/v3.28.7-fixed/python/test/test\\_apc\\_csv.py](https://github.com/OpenAPC/openapc-de/blob/v3.28.7-fixed/python/test/test_apc_csv.py) (accessed 13 September 2018).
32. OpenAPC results page at Travis CI:  
<https://travis-ci.org/OpenAPC/openapc-de> (accessed 13 September 2018).
33. Test results of a an OpenAPC build created on 23 June 2016:  
<https://travis-ci.org/OpenAPC/openapc-de/builds/139700030> (accessed 13 September 2018).
34. OpenAPC README file:  
<https://github.com/OpenAPC/openapc-de/blob/master/README.md> (accessed 13 September 2018).
35. OpenAPC README R Markdown template:  
<https://github.com/OpenAPC/openapc-de/blob/master/README.Rmd> (accessed 7 September 2018).
36. knitr: elegant, flexible, and fast dynamic report generation with R:  
<https://yihui.name/knitr/> (accessed 13 September 2018).
37. Marwick B, Boettiger C and Mullen L, Packaging Data Analytical Work Reproducibly Using R (and Friends), *The American Statistician*, 2018, 72(1), 80–88; DOI:  
<https://doi.org/10.1080/00031305.2017.1375986> (accessed 13 September 2018).
38. Datasets on fee-based Open Access Publishing:  
<https://openapc.github.io> (accessed 13 September 2018).
39. GitHub repository of the OpenAPC blog:  
<https://github.com/OpenAPC/openapc.github.io> (accessed 13 September 2018).
40. OpenAPC OLAP server:  
<https://olap.intact-project.org> (accessed 13 September 2018).
41. Cubes:  
<http://cubes.databrewery.org> (accessed 13 September 2018).
42. OLAP server, simple aggregate:  
<https://olap.intact-project.org/cube/openapc/aggregate> (accessed 13 September 2018).
43. OLAP server, drill-down operation:  
<https://olap.intact-project.org/cube/openapc/aggregate?drilldown=institution|period> (accessed 13 September 2018).
44. OLAP server, cut operation:  
<https://olap.intact-project.org/cube/openapc/aggregate?cut=institution:University%20of%20Cambridge|period:2016|publisher:Elsevier> (accessed 13 September 2018).
45. OpenAPC treemaps:  
<https://treemaps.intact-project.org> (accessed 13 September 2018).
46. OpenAPC data set, release version v3.28.7-fixed:  
<https://github.com/OpenAPC/openapc-de/releases/tag/v3.28.7-fixed> (accessed 13 September 2018).
47. Deinzer G, Wagner A, Broschinski C and Pieper D, Including Open Access Information in institutional repositories, 2017; DOI:  
<https://doi.org/10.5281/zenodo.887318> (accessed 3 September 2018).
48. ESAC – Joint Understanding of Offsetting, 2018:  
<http://esac-initiative.org/joint-understanding-of-offsetting/> (accessed 3 September 2018).
49. Kronman U, Evaluation of offset agreements – report 3: Springer Compact, *Technical report*, 2018, Swedish National Library:  
[http://openaccess.blogg.kb.se/files/2018/02/Evaluation\\_of\\_offset\\_agreements\\_SC\\_Report\\_3.pdf](http://openaccess.blogg.kb.se/files/2018/02/Evaluation_of_offset_agreements_SC_Report_3.pdf) (accessed 13 September 2018).
50. OpenAPC offsetting Dataset:  
<https://github.com/OpenAPC/openapc-de/tree/master/data/offsetting> (accessed 13 September 2018).
51. Offsetting OLAP cube, simple aggregate:  
<https://olap.intact-project.org/cube/openapc/aggregate> (accessed 13 September 2018).
52. Offsetting treemap:  
<https://treemaps.intact-project.org/apcdata/offsetting/> (accessed 13 September 2018).
53. Springer – Institutional Agreements, 2018:  
<https://www.springer.com/de/open-access/springer-open-choice/springer-compact> (accessed 3 September 2018).
54. Offsetting coverage OLAP cube, simple aggregate:  
[https://olap.intact-project.org/cube/offsetting\\_coverage/aggregate](https://olap.intact-project.org/cube/offsetting_coverage/aggregate) (accessed 13 September 2018).
55. Offsetting coverage treemap:  
<https://treemaps.intact-project.org/apcdata/offsetting-coverage/> (accessed 13 September 2018).
56. Broschinski C, Analysing the article coverage of offsetting contracts, INTACT, 2018:  
<https://www.intact-project.org/general/openapc/2018/03/22/offsetting-coverage/> (accessed 13 September 2018).
57. Wohlgemuth M, Rimmert C and Taubert N C, Publikationen in Gold-Open-Access-Journalen auf globaler und europäischer Ebene sowie in Forschungsorganisationen, Universität Bielefeld, 2017:  
<https://pub.uni-bielefeld.de/publication/2912807> (accessed 3 September 2018).

**Article copyright: © 2018 Dirk Pieper and Christoph Broschinski. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use and distribution provided the original author and source are credited.**



Corresponding author:

Christoph Broschinski

Software developer, OpenAPC and

Member, INTACT team

Universität Bielefeld

Bielefeld Nordrhein-Westfalen, DE

E-mail: [broschinski@uni-bielefeld.de](mailto:broschinski@uni-bielefeld.de)

Dirk Pieper

ORCID ID: <http://orcid.org/0000-0002-6083-9348>

To cite this article:

Pieper D and Broschinski C, OpenAPC: a contribution to a transparent and reproducible monitoring of fee-based open access publishing across institutions and nations, *Insights*, 2018, 31: 39, 1–18; DOI: <https://doi.org/10.1629/uksg.439>

Published by UKSG in association with Ubiquity Press on 10 October 2018