Article

Data-Driven Transition: Joint Reporting of Subscription Expenditure and Publication Costs

Irene Barbers [1,*,†] (ID)**, Nadja Kalinna** [1,2,†] (ID) **and Bernhard Mittermaier** [1,†] (ID)

1   Forschungszentrum Jülich, Central Library, D-52425 Jülich, Germany; n.kalinna@fz-juelich.de (N.K.);
    b.mittermaier@fz-juelich.de (B.M.)
2   RWTH Aachen, University Library, D-52062 Aachen, Germany
*   Correspondence: i.barbers@fz-juelich.de
†   These authors contributed equally to this work.

Abstract: The transition process from the subscription model to the open access model in the world of scholarly publishing brings a variety of challenges to libraries. Within this evolving landscape, the present article takes a focus on budget control for both subscription and publication expenditure with the opportunity to enable the shift from one to the other. To reach informed decisions with a solid base of data to be used in negotiations with publishers, the diverse already-existing systems for managing publications costs and for managing journal subscriptions have to be adapted to allow comprehensive reporting on publication expenditure and subscription expenditure. In the case presented here, two separate systems are described and the establishment of joint reporting covering both these systems is introduced. Some of the results of joint reporting are presented as an example of how such a comprehensive monitoring can support management decisions and negotiations. On a larger scale, the establishment of the National Open Access Monitor in Germany is introduced, bringing together a diverse range of data from several already-existing systems, including, among others, holdings information, usage data, and data on publication fees. This system will enable libraries to access all relevant data with a single user interface.

Keywords: open access; APC; workflow; journal subscription; offsetting; publication fee; monitoring; transition

1. Introduction

The open access publishing model is an inherent part of scientific publishing today, and needless to say it affects the long-term tasks of academic libraries in procuring scholarly content. This content is no longer solely available through the traditional model of journal subscriptions, but increasingly in the form of accessible online content that is free of charge. This relatively new area of responsibility that falls upon libraries includes supporting affiliated scientists with the publishing process, as well as offering workflows and infrastructure such as institutional repositories or publication funds to handle article processing charge (APC)-financed open access publications. Many libraries manage publication funds to cover these costs and therefore have to report on this expenditure. Although this role has been established recently with the growing acceptance of and demand for open access publishing, there are still challenges associated with the large-scale transition to open access. One of these challenges is the reconciliation of parallel expenditure for subscription fees for closed access content and the rising costs for publishing in open access journals, which at some point in the future could make subscription costs obsolete. Libraries should record data for both types of expenditure and they need to implement joint reporting to manage the shift from subscription budget to publication budget calculations. Section 2 will illustrate why it is important to develop workflows to collect data on publication charges and how

libraries should be involved in this process. Section 3 then presents a method to create a joint reporting of subscription costs and publication costs through adapting two separate systems to enable combined queries. Section 4 then shows examples for results and discusses the use of joint reporting. Further to this, in Section 5 we provide an outlook on the National Open Access Monitor, a reporting system in Germany at the national level.

2. Workflows for Collecting and Using Data on Publication Charges

In this section, the relatively new roles of libraries regarding open access publishing will be presented. One of them is a comprehensive service in the field of scientific publishing, which comprises the development of workflows for collecting and using data on publication charges. The collection and usage of data on publication charges aim to address the need for reliable expenditure data, not only when it comes to budget plans and negotiations with publishers for new agreements like "read and publish" contracts (combining subscription with open access publishing) for the individual research institution, but also for national agreement negotiations and overviews of the costs of scientific publishing. As open access publishing increasingly becomes a requirement, in particular for the APC-funded business model, the question arises as to who is responsible for managing the growing number of publication fee payment processes. Originally, these fees were charged to the publication's authors and were known as "author fees". However, there is a high level of administrative effort involved and authors realized that when they processed the APC payment themselves, it often took them more than one hour. This was revealed in a survey for a Knowledge Exchange report in 2017, in which authors said that "streamlined administrative procedures" would be of great help, and they would make expenditure become more transparent [1] (p. 14). Another UK study about the total costs of open access inferred that the management of gold open access could take two hours per article [2] (p. 2 and p. 17). A number of libraries therefore started to process these payments as a service provided in the field of scientific publishing. However, the question regarding who is currently responsible for the management of publication fees is not easy to answer, since it is handled in various ways—even within one country with different scientific institutions. A study by Solomon and Björk to estimate per-article expenditure for the publications of research-intensive universities in the USA and Canada indicates that there was no reliable data on publication fees collected in the USA and Canada (publication data collected between 2009 and 2013), since they used data from the APC payment repositories of four European institutions [3] (p. 5). In the UK, Wellcome Trust and Research Councils UK (RCUK) cover the majority of APCs for publications that contain publicly-funded research results. Pinfield et al. (2015) examined the payment responsibilities of 23 universities in the UK and stated that the APCs of these institution were mostly financed by the two aforementioned organizations [4] (pp. 1757–1758). In addition, Pinfield/Finch (2014) reported that the institutions' libraries often manage these payments centrally (ibid., p. 1752). In German universities and non-university research institutions it is more common practice for affiliated libraries assist their publishing scientists, particularly when they choose open access journals [5] (p. 87). The German Research Foundation (DFG) has, similarly to the Wellcome Trust or RCUK, established the program for research infrastructure known as "Scientific Library Services and Information Systems (LIS)", which comprises DFG funding for open access. German universities can apply for these funds, which "are intended as start-up funding to set up an open access publication fund [...]" [6] (p. 3). It is mostly the university libraries that are responsible for managing these open access funds, for example the Leibniz Universität Hannover (TIB) [7].

However, thinking of ways to finance through external funding options, some argue that libraries cannot simply be declared as the responsible institutions that centrally take care of publication charge payments. There are organizational issues due to structural conditions such as the two-tier library system in Germany. In such cases, the separation of a library system into a central library and several independent institute libraries affects the assignment of university-wide tasks of the university library. Nevertheless, the central management of payment by the libraries is highly desirable. Libraries have

supplied scientific literature from closed access publishers for many years. The nature of this library service is currently changing, however, with at least one-third of scholarly journal literature being freely available due to open access publishing [1] (p. 14). Giving the responsibility of managing publication payments to libraries is a good idea, since, as mentioned before, they have been negotiating subscription contracts with publishers for years. This is a great advantage when it comes to the reconciliation of parallel expenditure for subscription fees because they have the financial data available. Furthermore, they often report on the publications output (bibliography) of their research institutions, meaning they are already involved in the field of scientific publishing. Non-university research institutes often run their own publication funds and have therefore implemented open access policies to communicate the support and services they offer in terms of open access publishing and strategies for approaching the transition. We, the Central Library of Forschungszentrum Jülich, are a non-university research library with over 200 publications per year for which invoices are processed, amounting to roughly 400,000 EUR in APCs, hybrid publication charges, and associated costs (see Section 3). According to our Open Access Strategy, "the Central Library has the task of supporting the transition from subscription journals to open access and to control expenditures for subscription journals in such a manner that sufficient funds are available for gold open access publication fees (article processing charges, APCs)" [8]. This claim is compliant with the OA2020 initiative's mission statement, that in order to be successful in the transition process it is necessary to "establish transparency with regard to costs and potential savings" [9]. Therefore, collected data can be used to prepare for negotiations with publishers when it comes to discussing offsetting publication costs against subscription costs or of flipping journals from the hybrid model to the full open access model. To summarize, it should not fall to the authors to contend with APC management. The libraries have experience managing subscription agreements and budget plans, and also often take care of external funding budgets (DFG, RCUK, etc.) and bibliographies. The library should therefore be responsible for processing all publication fee invoices, including page or colour charges, and invoices for article processing charges (APCs) for publishing in full or hybrid open access journals. If this is not implemented, there will be little opportunity to monitor the overall costs of subscription and publishing [10] (p. 325). However, there is a need for optimization articulated even in cases where libraries such as the German Max Planck Digital Library are in charge of payment management. Responsible staff at the Max Planck Digital library for example remarked that in contrast to the proven structures in managing journal subscriptions "the management of publication costs is still largely based on manual processes" [5] (p. 87). They also mentioned the great need for standardization in this area, as articulated by the British Research Information Network (RIN) (ibid.). The RIN did not publish any new findings in this regard on its website in the meantime [11]. Despite that, the MDPL described that the expenses for licenses, subscriptions and publication fees should be brought together to be able to evaluate publishers' offers, but they do not explain how they are achieving that reconciliation of parallel expenditure [5] (p. 88). In accordance with these statements, Vierkant et al. outlined that it is currently hardly possible to automate the APC workflows [12] (p. 164). This means that publication fee management must be performed manually, which in turn implies that the library must cooperate closely with the authors. If an author is planning a submission, he/she should inform the library in good time. When the library or the research institution as a whole has distributed an open access strategy/policy in which funds and prepayment methods are also communicated, authors tend to get in contact with the library on time for payment reasons. This enables the library to check the corresponding author's affiliation, the business model offered by the journal, and the total applicable publication fees. The library also clarifies if there is a cooperation contract with the publisher[1]. These conditions affect the fund from which the publication can be paid for and the payment process itself, as these agreements often imply other payment methods, such as payment from a prepayment account to reduce single invoices

---

[1]　　For example the publishing agreement between Frontiers and the Helmholtz Association.

and a reduction in publication fees by receiving discounts granted in such cooperation agreements. This emphasizes the complexity of the APC market, which must be considered when establishing workflows and systems to collect and use publication fee data [3] (p. 7). The MDPL also brings up this subject by commenting upon the hurdle of intensive dialogue with each individual publisher/provider which must take place, as the administrative process varies from provider to provider [5] (p. 88). Therefore librarians must be familiar with the changing market of scientific publishing and ways of distributing published research results. Keller described the growing need for competences in this field with the term "publishing literacy", which comprises "the in-depth understanding of new (digital) possibilities of scholarly communication and the skills of the individual researcher to select the best format to present, publish and disseminate his results and ideas" [13] (p. 158). The author describes this skill as an important aspect of information literacy and as a "key focus of the work of academic librarians" in supporting the author adequately (ibid.). This task includes checking for green open access options offered by publishers and implementing the procedures that automatically make the permitted manuscript version available in the library's repository after the applicable embargo period has elapsed. However, the above-mentioned good practice requirements give rise to the question as to whether there is a suitable place and a proven workflow for storing and reporting on data on publication fees in academic libraries. Vierkant et al. note that there is no established approach so far [12] (p. 158). The majority of universities and research libraries run online repositories with metadata on the institution's publications. However, these repositories do not provide a means of managing invoicing data by default. Nevertheless, some academic libraries, such as the University of Regensburg, have started to implement APC management within their repositories [14]. In order to create a fast and practical way of handling publication fees, we initially set up a database using the software MS Access. We called this the Publication charge-Reprint-Colour Charge-Cover (PROCC) database. The following data is entered into this database for each publication that incurs charges:

- Title
- Source (name of journal or e-book)
- DOI
- Publisher
- Corresponding author
- Institute
- Status of payment, cost type (e.g., permission, APC, colour charges), price, invoice, and date processed
- Special contract (prepayment account, offsetting)

The PROCC database comprises a simple, user-friendly interface where the data can be easily inputted. Data storage and maintenance is server-sided (see Section 3). The database can be easily queried to determine the total expenditure, for example, per publisher, per year, or per cost type. However, running a repository for all publications by our institution implies that we have to maintain two separate databases, which in turn means that there are data overlaps because metadata such as title, author, and journal is recorded both in the repository and in the PROCC database. In other words, the database containing publication costs exists separately from the repository's record of the publication. They are only connected by the ID number of the related record, which has to be entered manually. As Wagner pointed out, the publishing process should be changed to avoid multiple records of the same items of data [15] (p. 43). Such time-consuming workflows should be optimized by using established systems. A clear disadvantage of our separate database is the manual data input, which leads to duplication (title, author, source, institute, etc.) which is moreover non-transparent because the expenditure is not accessible by the scientific institutes. For these reasons, a decision has been made to shut down the PROCC database. Comparable to the plans of the University of Regensburg and in cooperation with Deutsches Elektronensynchrotron (Hamburg/Zeuthen) (abbreviated as DESY), we have expanded our shared repository infrastructure [16] to include the functionality of recording publication fees. In this context, an Extensible Markup Language (XML)-based format including cost information in the Open Archives Initiative—Protocol for Metadata

Harvesting (OAI-PMH) interface has been developed in order to be harvested by the University Library of Bielefeld, which releases datasets for the Open APC Initiative [17]. At the moment, we still store the data on fees in both databases (the PROCC database and the APC management tool in the repository) and we will continue to do so until we have a solution for connecting the repository with the electronic resource management system (ERMS). Until then, we can exploit the advantages of the SQL-based PROCC database, which enables joint reporting with the expenses for subscription costs as it shares the same SQL server as the ERMS.

3. Electronic Resource Management System and Publications Database—Joint Reporting System

During our search for methods of gathering data for joint reporting, we found that such methods are mostly depicted in a general way, for example "gathering data from volunteer libraries" to calculate the "total cost of ownership", as described by the authors of [4] or [18]. However, where and how institutions collect or even combine both kinds of data at a local level has not been reported. This lack of information shows the need to share examples of methods, which is what we are currently attempting to achieve.

Our requirements for the joint reporting system include the immediate availability of data with subsequent daily updates. The use of our established two local systems has proven to be an obvious and practical approach. Data on subscription journals and on publication fees are updated in both systems in an ongoing daily workflow and can be used for real-time reporting. Additional advantages can be seen in the use of skills developed for the established workflows, the continuation of the established workflows themselves, and the avoidance of duplicate data storage.

Having described the process and workflows for recording data on publication expenditure in the previous section, the following section will provide details on how subscription costs are captured with the help of an electronic resource management system (ERMS) and how automated joint reporting of expenditure for subscriptions and publications has been established. An ERMS is defined by Verminsky and Blanchat as: "an internal system or software program to assist in the maintenance of licensed electronic resources, such as databases, ebooks, and ejournals. An ERMS may include a means to track license agreements including license and copyright terms, renewals, access management, and collection development" [19] (p. 235). ERMSs are in most cases not part of library management systems [20], but stand alone, which makes the integration of electronic resource management into a library's workflows challenging. In addition, ERMSs do not provide the technical or organizational framework to handle a publication fee workflow.

Our institution operates its own ERMS [21] which is built as an SQL Server database. Development started in 2007 [22], shortly after the library switched its subscription model from print to e-only wherever possible. The ERMS provides information about journal metadata and access information, imported from the "Elektronische Zeitschriftenbibliothek" ("Electronic Journals Library"), hosted by the University Library of Regensburg [23]. Certain licensing information such as subscription costs, cost codes, and licensing codes are not managed in the EZB database, but are stored and managed inside the local ERMS. COUNTER usage reports [24] are collected on a regular basis from publishers' platforms or their usage report providers, and uploaded into the ERMS. Licensing information and usage data are then used to provide reporting in several views, tailored to the needs of collection development. These views present reports detailing subscription costs, usage data, and cost per download analysis for each publisher as well as at the individual journal title level. This includes in-depth analysis of package subscriptions, time-lines, and usage data broken down by year of publication as well as usage reports for open access publishers and open access journals.

One step towards the goal of supporting the shift away from subscription budgets to open access publication budgets is to monitor both budgets closely in order to be able to track developments. We therefore implemented several reports into our ERMS monitoring web interface, showing the number of publications by journal and the publication costs at the journal level and publisher level. The only source for these reports is the PROCC database, as described in the previous section. To go one

step further in bringing both sides together, SQL statements across the two databases provide reports which unite subscription costs, downloads, cost per download, and publication costs per journal into one single view. Both databases run on the same SQL Server and reports are displayed within one single web interface built with SQL Server Reporting Services. The main obstacle when creating joint reports is that the ERMS presents subscription information at a different level than the PROCC database, which presents information on publications. On the one side, we have information about licensing at the journal level, package level, or publisher level, whereas on the other side, data on publication expenditure is presented at the article level. To align these different levels, publication costs must be aggregated first at the journal level. This is achieved by initially creating a view of publication costs at the journal level which is then used to serve as a source for the combined query to join subscription data and publication data. Additionally, a common identifier for journal titles is necessary in both databases, and as the identifier of the EZB, termed EZB-ID, is already in use in the ERMS, the same identifier was added to the journal table in the PROCC database. Another possible identifier would be the identifier used in the "Zeitschriftendatenbank – ZDB" [25]. The use of ISSN has been proven unsuitable as there are too many inconsistencies in this identifier. Aggregating reports at the publisher level is an additional challenge. As there exist no authorized forms for publishers in the commonly used knowledge bases, publisher names used in ERMS are often different to publisher names used in the PROCC database. Again, we regularly find different levels of aggregation and allocation of journals. In the past, we used to allocate journals to hosts rather than publishers in the ERMS, as usage reports come from the hosts. Publishers and hosts often differ. Highwire [26] is a good example. It is an organization that serves as a host for multiple publishers and societies. All journals hosted by Highwire used to be joined under the same label in ERMS in the past. However, joint reporting with publication expenditure means that this allocation of journals no longer meets the purpose of reporting. To match publishers in both ERMS and PROCC databases, several adjustments had to be made in both databases. As a first step, the bigger hosts in ERMS, where several publishers and societies had been summarized together, have been split up so that those publishers are now represented individually. The second step was to insert the data on ERMS publishers into the PROCC database publishers table to match the publisher names and thus to be able to join the respective publishers tables in SQL statements for the creation of reports.

As we are building on established structures of the ERMS that were originally designed for the sole purpose of usage reporting, some other adjustments were necessary. In the past, reporting was only performed for those journals where usage data was provided and where subscription costs had been incurred. The philosophy had always been that the usage data of journals with no subscription costs was not of central interest, as there was no cost per download as such to calculate. This also meant that initially whenever a journal had no usage reported and/or no subscription expenditure recorded, it did not show up at all, even if publication charges had been incurred. In other words, if a certain journal was not already captured by ERMS, it would not show up in the joint reporting interface, even if publication fees had been allocated to it in the PROCC database. One measure in opening up ERMS structures has therefore been to input open access journals from well-known publishers manually on a regular basis in ERMS. This can be done by exporting so-called "green" journals from a chosen set of publishers from our knowledge base, the EZB. Another idea would be to harvest selected data from the Directory of Open Access Journals (DOAJ) [27]. However, an automated harvesting process for relevant open access journals is still desirable.

4. Results and Further Use of Joint Reporting

One of the major benefits of joint reporting is that the complete range of a given publisher's portfolio, be it of subscription journals, hybrid, or open access journals, can be viewed and analysed in a single report with all the relevant data aggregated. The joint view thus provides us with new key figures like total expenditure per journal and total expenditure per publisher. Factoring publication fees into the calculation broadens the scope and brings a fresh perspective to the traditional method

of collection development through cost-per-download analysis. Additionally, the automated joint reporting serves as the basis for further analysis and visualization of the institutional expenditure on subscriptions and publications as a whole and with a focus on the most important publishers. The visualization is presented as the Open Access Barometer [28] on the institution's website and constitutes an example of the visibility and further use of the data. The purpose of presenting the information on the website is to advocate open access within our institution and to demonstrate the library's efforts towards supporting our institution's open access goals through the considered collection development. The total expenditure, number of the institution's own publications, cost per publication, and distribution of expense types are documented for the 12 most important publishing houses (according to expenses and number of publications). Comments on the cost structure and the transition status of the individual publishers are additionally exclusively available on the intranet. Figure 1 shows the shares in the total number of publications of the most important publishing houses (according to expenses and number of publications) where journal articles with a corresponding author from our own institution were published in 2017, and shows the shares of total expenditure for those same 12 publishers. "Total expenditure" adds the sum of subscription fees and publication fees paid to the publishers in 2017.
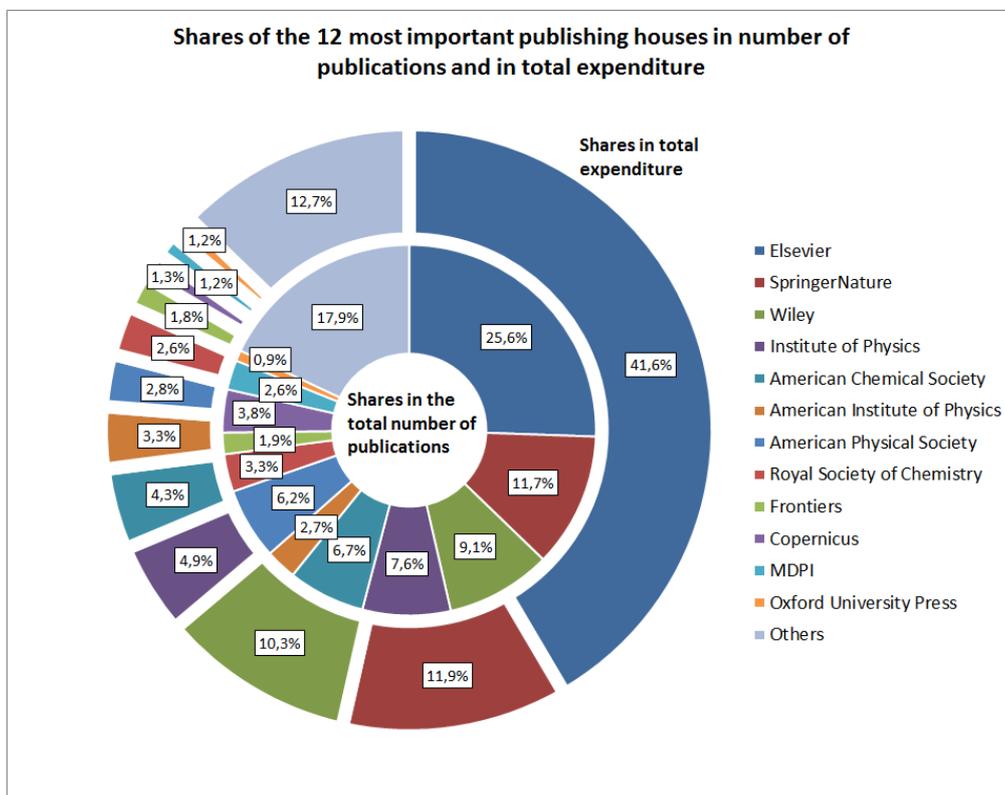

Figure 1. Shares in the total number of publications and in total expenditure in 2017. Source: Forschungszentrum Jülich.

The three biggest publishers with the highest shares in the number of publications (Figure 1) together account for access to 62.3% of all subscribed journals and for 70% of subscription expenditure Table 1. The table shows the shares of subscribed journals compared to the shares of expenditure for the nine subscription publishers from our previous sample.

Table 1. Shares of the number of subscribed journals and of subscription expenditure.

Publisher	Share of the Number of Subscribed Journals	Share of Subscription Expenditure
Elsevier	51.5%	50.3%
- Elsevier Subscribed Journals	2.4%	39.7%
- Elsevier Freedom Collection	49.1%	10.6%
Springer Nature	9.0%	9.1%
- Springer Subscribed Journals	0.8%	3.7%
- Springer Cross Access	5.8%	0.4%
- Nature Subscribed Journals	0.4 %	4.3%
- Nature Complete	1.9%	0.7%
Wiley	1.8%	10.6%
Institute of Physics	1.8%	4.7%
American Chemical Society	1.4%	5.2%
American Institute of Physics	0.4%	3.2%
American Physical Society	0.3%	2.6%
Royal Society of Chemistry	1.0%	2.7%
Oxford University Press	3.6%	1.0%
Others	29.5%	10.7%
Total for all publishers	100%	100%

Figure 2 presents the proportion of the types of cost for each publisher. The chart shows that hybrid publication charges and associated costs are more significant than gold open access expenditure for the traditional subscription publishers. The exception here is SpringerNature, which has a large proportion of APC expenditure due to a high number of publications in both Nature Communications and Scientific Reports.
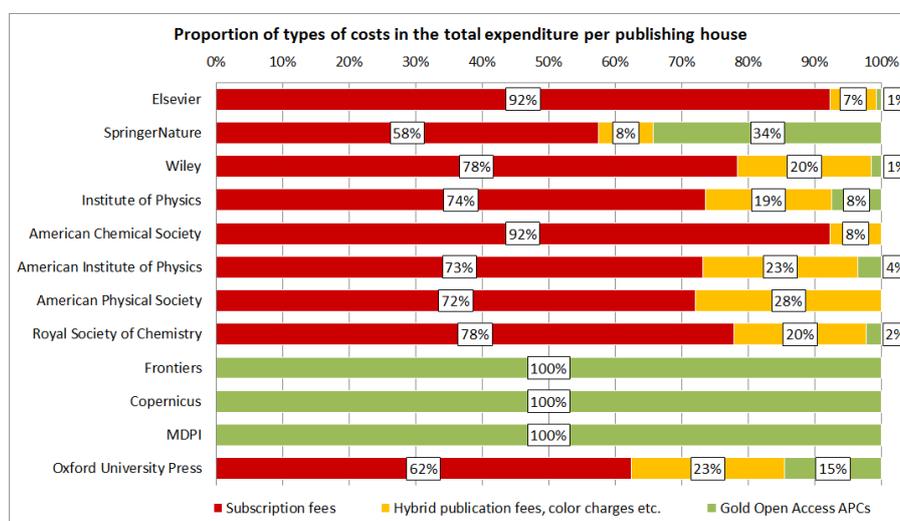
Figure 2. Proportion of types of costs in 2017. Source: Forschungszentrum Jülich.

For almost all the publishers, we can see that their average hybrid publication charges are higher than their average gold open access APC Table 2. This corresponds with the findings of a Universities UK report [29] that shows the difference in APC price bands for full open access journals and hybrid journals.

Table 2. Average publication expenditure by type of cost.

Publisher	Average APC in EUR	Average Hybrid Publication Charge in EUR	Average Color Charge in EUR	Average Cover Charge in EUR	Average Page Charge in EUR
Elsevier	1017	2092	1049		925
Springer Nature	1840	2634	1763		
Wiley	1017	2416	1198	1242	1924
Institute of Physics	1346	1901	1565		405
American Chemical Society		1127			
American Institute of Physics	2042	1950			473
American Physical Society		1269	1643		
Royal Society of Chemistry	516	1659		1117	
Frontiers	1650				
Copernicus	1447				
MDPI	1049				
Oxford University Press	2130	2450	1050		900
Others	1209	1052	1113		1252

The figures shown in Tables 2 and 3 are based on the publication fees that were processed by our library in 2017. It is worth noting that the number of instances for APCs and other charges (Table 3) does not correspond to the number of journal articles with corresponding authors from our institution used for Figure 1. This is partly due to the corresponding author articles being published in subscription journals, where no publication fees are incurred. Furthermore, publication fees can be paid for by institutions other than our own and some authors do not use the library's publication workflow services. It should also be noted that fees can vary for a single publisher depending on the journal selected for publication.

Table 3. Total instances and total sum of publication fees.

	APCs	Hybrid Publication Charges	Color Charges	Cover Charges	Page Charges
Total instances	138	69	33	6	28
Total sum of fees	209,064 EUR	117,166 EUR	39,038 EUR	7076 EUR	33,397 EUR
Total average publication fee	1510 EUR	1698 EUR	1183 EUR	1179 EUR	1193 EUR

What use can we make of these findings? The joint reporting enables control over expenditure in a manner that places us in a position to make fully informed decisions on collection development in

the light of the transition process. Download metrics and the calculation of cost per download are still of significance for the detailed collection development of subscription journals, since these metrics are informative for cancellation and/or renewal decisions at the single-journal or package levels. However, data on the types of cost, shares of subscription expenditure, and publishing fee rates help to create an overall picture, which together with the key figure of "total expenditure per publisher" will indicate where attention should be focused in terms of negotiation strategy. Over the course of the next few years, we will be able to create views on the development of expenditure and keep track on shifts in proportion to demonstrate the progress of transition and collected data can be used to prepare for negotiations with publishers.

5. The Broader Perspective: A National Open Access Monitor for Germany

The system described in the previous sections can be adopted by any academic library. In addition to some programming skills, it is necessary to

- maintain an ERMS with an appropriate interface
- download COUNTER statistics of subscribed journals and store them in a database
- maintain the bibliography of the institution with authority data for journals and publishers
- record the open access and non-open access publications fees paid by the institution, again using authority data for journals and publishers

However, in most cases not all requirements are met, which means that the number of libraries utilizing such a system is quite limited. Furthermore, it would be rather uneconomical if all libraries were to develop and implement such systems independently. This was one of the reasons behind establishing the National Open Access Monitor in Germany. It originated within the framework of the OA2020 movement, where the German Alliance of Research Organisations [29] established a National Contact Point [30] in 2017. One of the tasks of this National Contact Point is to establish a data monitor [31]. This data monitor, which is still under construction, received a powerful boost by a grant from the Federal Ministry of Education and Research (BMBF). The BMBF had established a National Open Access Strategy for Germany in autumn 2016, which included the idea of the National Open Access Monitor. In a subsequent call, the Central Library of Forschungszentrum Jülich was the successful applicant for this task; the project started in January 2018. According to the National Open Access Strategy [32], the Open Access Monitor should "*track the quantitative status of Open Access in Germany in a reliable manner. (...) If institutions are able to quantify the share of Open Access in their publications, they will also be able to identify areas that are weaker in Open Access and foster Open Access in a more targeted manner. It is also planned that monitoring will show what sources and amounts of funds are used for obtaining scientific information and for financing publications (both for Open Access and the subscription-based model). In this way, the changeover to Open Access can be structured in a tailored manner.*" This means that the Open Access Monitor has a broad focus—not only open access, but also subscription, and not only a from national perspective but also with the local view of single institutions. Academic publishing is currently a mixed economy, consisting both of the open access model and the subscription model, and it will remain so in the foreseeable future. Therefore, data from both models has to be gathered. Some of the data is available only at the local level and therefore must be collected there. This data includes expenditure (both for APCs and subscriptions), journal holdings, and usage data. Some of the data are already available in databases and therefore does need not to be collected at the local level. This includes publication and citation data. The National Open Access Monitor does not intend to collect the data directly from the participating institutions nor to build a new system comprising all necessary functions to do so. In contrast, for reasons of economy and acceptance in the community, the Monitor will rely on systems which already exist or which are currently being developed (Table 4).

Table 4. Sources for the National Open Access Monitor.

System	Data	Operator
Global Open Knowledgebase (GOKb) [33]	Authority records for journals and publishers	Staatsbibliothek zu Berlin and VZG
Virtual International Authority File (VIAF) [34]	Authority records for institutions	OCLC
LAS:eR [35]	Licenses and license fees	hbz, VZG, University Libraries of Frankfurt and Freiburg
National Statistics Server [36]	COUNTER statistics	University Libraries of Frankfurt and Freiburg
KB-database [37]	Publications and citations	Kompetenzzentrum Bibliometrie
BASE [38]	Publications	University Library Bielefeld
OpenCitations [39]	Citations	David Shotton and Silvio Peroni
Open APC	Publication fees	UB Bielefeld

License information will come from LAS:eR, a system currently being developed within the framework of a DFG project. It is intended to be a free and open ERMS, avoiding the many problems associated with commercial systems on the market. Today, the extent to which libraries will use this system is unclear. However, we are optimistic that there will be reasonable uptake: it is intended to establish a private Lots Of Copies Keep Stuff Safe (LOCKKS) network at a national level [40], where the licensing rights of the respective members will be maintained in LAS:eR. Usage information will be incorporated as COUNTER statistics. An easy option is participation in the National Statistics Server, where institutions can maintain their usage statistics; on their side, only the Standardized Usage Statistics Harvesting Initiative (SUSHI) [41] credentials have to be provided. Institutions that do not use this fee-based service could provide their COUNTER statistics directly. For publication and citation data, a number of sources[2] are used because no single source fits all purposes: BASE contains publications from all journals, but not from all institutions (and from some institutions not covering the complete bibliography), because institutions have to participate actively. The Web of Science [42] comprises both publication and citation information, but not for all journals. No interaction on behalf of the institutions is necessary, so the coverage for the indexed journals is very good. The Web of Science will be accessed via the bibliometric database of Kompetenzzentrum Bibliometrie, where the affiliations of German institutions have already been normalized and cleaned. The idea is to calculate ratios between the number of publications in the Web of Science and the number of publications in BASE at the level of publishers for a number of institutions with an excellent coverage in BASE. These ratios can then be used to extrapolate for other institutions. OpenCitations is freely accessible and contains citation data from most publishers. As of January 2018, notable exceptions are Elsevier and the American Chemical Society. Publication fees can be taken from Open APC. Here, an increase in participation of institutions and a broader coverage of types of publication fees is desirable. All systems

---

2   Additionally, the use of Crossref metadata is considered.

share a common feature in that they do not contain data at a 100% level; in some cases, the figure is much lower. Therefore, it will be an important goal of the project to identify and develop means of increasing the participation of institutions in collecting data and delivering it to the respective systems. A key problem is authority control: some systems do not have validated authority records[3]; while others implement authority control but utilize their own, proprietary scheme[4]. To map data from different sources, it is therefore necessary to validate different proprietary schemes against a master file of authority records. For institutions, we use the Virtual International Authority File (VIAF), and for journals and publishers the Global Open Knowledgebase (GOKb). GOKb will be a machine-readable excerpt of the "Zeitschriftendatenbank—ZDB", the worlds' largest database for serials. GOKb already exists, but will undergo a fundamental update in the near future. The National Open Access Monitor will amalgamate this data in a PostgreSQL 10 database. It will come with a user interface that allows access to the data. A dedicated rights management for the regulation of access to the database in various usage scenarios will be implemented:

- Institutions will have access to their own data from a single user interface, in most cases for the first time.
- Negotiators for transformation agreements will have all relevant data concerning a specific publisher at hand.
- On behalf of the BMBF, the status and progress of open access can be monitored on a national scale by Forschungszentrum Jülich or other parties.
- Scientists can use the data for research on the publishing system.

6. Conclusions and Outlook

Open access has brought a change in the roles and systems of academic libraries. A dedicated open access strategy is desired, funds have to be established and maintained, and a central payment process managed by the library is desirable. This will allow publication cost data to be collected and stored in an appropriate system, which should be connected to the institutions' bibliography/repository and augmented with subscription expenditure information. If a library currently has only a bibliography or a repository in place, it is worth considering expanding their system's capacity to cover publication fees instead of creating a new system in parallel. This system, of course, still has to be connected to an ERMS. The collection of this data is not an end to itself. The information helps the library in its reporting duties and in dealing with questions of budget allocation. It enables librarians to engage in negotiations with publishers about offsetting publication costs against subscription costs or flipping journals from the hybrid model to the full open access model. Since we suggest that offsetting should only be seen as an intermediate step within the transition process [43], the importance of detailed expenditure control cannot be stressed enough. Negotiations can only be pursued on the basis of solid data. Offsetting agreements must have valid data on a larger scale at hand, potentially at a national level. Without such data, endeavours such as the DEAL negotiations [44] cannot be pursued. At a minimum, the current expenditure levels and the publication numbers must be at hand. To this end, anticipatory aggregation at the national level is convenient and economical. Furthermore, this approach reduces the burden on individual institutions to maintain such a system on their own. To some extent, however, the collection of data will still reside with each institution.

Important conclusions we can draw from our experiences:

- It is important to use normalized data/authority files. Otherwise, reporting at any intermediate level (e.g., the publisher's level) involves tedious manual work. Without controlled data, ease of reporting is only feasible at a small-scale level (e.g., single articles) or at the top level.

---

3   Example: Names of journals and publishers in BASE.
4   Example: Names of journals and publishers in Web of Science.

- It makes sense to use one single system that incorporates both (open access) publication information and subscription information.
- At a national level, data should be available in stock and not collected on demand. The latter involves multiple requests for the delivery of data, sometimes on short notice, and therefore rather poor compliance and a higher workload for the local libraries.

Author Contributions: Irene Barbers is responsible for Sections 3 and 4; Nadja Kalinna is responsible for Sections 1 and 2; Bernhard Mittermaier is responsible for Sections 5 and 6.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

APC	Article Processing charges
ERMS	Electronic Resource Management System
OAI-PMH	Open Archives Initiative—Protocol for Metadata Harvesting
PROCC database	Publication charge-Reprint-Colour Charge-Cover database
SUSHI	Standardized Usage Statistics Harvesting Initiative
XML	Extensible Markup Language
LOCKKS	Lots Of Copies Keep Stuff Safe

References

1. Knowledge Exchange: Paying for Open Access. The Author's Perspective 2017. Available online: http://doi.org/10.5281/zenodo.438037 (accessed on 17 April 2018).
2. Research Consulting UK: Counting the Costs of Open Access. The estimated Cost to UK Research Organisations of Achieving Compliance with Open Access Mandates in 2013/14, 2014. Available online: http://www.researchconsulting.co.uk/wp-content/uploads/2014/11/Research-Consulting-Counting-the-Costs-of-OA-Final.pdf (accessed on 17 April 2018).
3. Solomon, D.; Björk, B.-C. Article processing charges for open access publication—The situation for research intensive universities in the USA and Canada. *PeerJ* **2016**, *4*, e2264.
4. Pinfield, S. The "Total Cost of Publication" in a Hybrid Open-Access Environment: Institutional Approaches to Funding Journal Article-Processing Charges in Combination With Subscriptions. *J. Assoc. Inf. Sci. Technol.* **2016**, *67*, 1757–1766.
5. Sikora, A.; Geschuhn, K. Management of article processing charges—Challenges for libraries. *Insights* **2015**, *28*, 87–92.
6. Deutsche Forschungsgemeinschaft. Guidelines Open Access Publishing. DFG form 12.20—02/17. Available online: http://www.dfg.de/formulare/12_20/12_20_en.pdf (accessed on 13 February 2018).
7. Leibniz Information Center for Science and Technology University Library: Leibniz Universität Hannover's Open Access Publishing Fund. Available online: https://www.tib.eu/en/publishing-archiving/open-access/financing-open-access/publishing-fund-leibniz-universitaet/ (accessed on 13 February 2018).
8. Forschungszentrum Jülich. Open Access Strategy of Forschungszentrum Jülich. Available online: http://www.fz-juelich.de/zb/EN/Expertise/open_access/oa_strategy_fzj/oa_strategy_node.html (accessed on 21 April 2018).
9. OA2020, Mission Statement. Available online: https://oa2020.org/mission/ (accessed on 9 January 2018).
10. Frick, C. Empfehlungen für Workflows zur Übernahme von Publikationsgebühren. In *Praxishandbuch Open Access*; Söllner, K., Mittermaier, B., Eds.; De Gruyter Saur: Berlin, Germany; Boston MA, USA, 2017; pp. 323–330, ISBN 978-3-11-049406-8. Available online: https://www.degruyter.com/downloadpdf/books/9783110494068/9783110494068-037/9783110494068-037.pdf (accessed on 30 January 2018).
11. The Association of Commonwealth Universities. Research Information Network: APC Payments. Available online: https://www.acu.ac.uk/research-information-network/apc-payments (accessed on 30 January 2018).

12. Vierkant, P.; Siegert, O.; Deinzer, G.; Gebert, A.; Herbstritt, M.; Pampel, H.; Tobias, R.; Wagner, A. Workflows zur Bereitstellung von Zeitschriftenartikeln auf Open-Access-Repositorien—Herausforderungen und Lösungsansätze. *Das Offene Bibliotheksjournal* **2017**, *4*, 151–169.

13. Keller, A. Publikationskompetenz als neues Aufgabengebiet für Bibliotheken: Eine australische Fallstudie. *Bibliothek Forschung Praxis* **2015**, *39*, 158.

14. Deinzer, G. Managing APCs with institutional repository software (EPrints). In Proceedings of the 11th International Conference on Open Repositories, Trinity College, Dublin, Ireland, 13–16 June 2016. Available online: https://epub.uni-regensburg.de/35536/ (accessed on 30 January 2018).

15. Wagner, A. Publizieren ist Nicht Genug. In *Der Schritt Zurück als Schritt Nach Vorn—Macht der Siegeszug des Open Access Bibliotheken Arbeitslos? In Proceedings of the 7. Konferenz der Zentralbibliothek Forschungszentrum Jülich WissKom 2016, Jülich, Germany, 14–16 June 2016*; Mittermaier, B., Ed.; Forschungszentrum Jülich GmbH Zentralbibliothek, Verlag: Jülich, Germany, 2016; pp. 23-45, ISBN 978-3-95806-146-0.

16. JOIN²—Just anOther INvenio INstance. Available online: https://join2-wiki.gsi.de/cgi-bin/view (accessed on 16 January 2018).

17. INTACT Open APC. Available online: https://www.intact-project.org/openapc/ (accessed on 20 December 2017).

18. Shamash, K. Article Processing Charges (APCs) and Subscriptions: Monitoring Open Access Costs 2016. Available online: https://www.jisc.ac.uk/reports/apcs-and-subscriptions (accessed on 18 March 2018).

19. Verminsky, K.A.; Blanchat, A. *Fundamentals of Electronic Resources Management*; Facet Publishing: London, UK, 2017; ISBN 9781783302307.

20. Kowalak, M.; Sabisch, A. Electronic Resource Management. In *Neue Formen der Erwerbung*; Göttker, S., Wein, F., Eds.; De-Gruyter Saur: Berlin, Germany, 2013; p. 150, ISBN 9783110255508.

21. Heinen, I. Das Jülicher Electronic Resource Management System. *B.I.T. Online* **2011**, *14*, 41–43.

22. Aumeier, F.; Heinen, I. Inhouse-Lösung für das Jülicher Electronic Resource Management-System. *Bibliotheksdienst* **2007**, *41*, 322–330.

23. Elektronische Zeitschriftenbibliothek. Available online: https://ezb.uni-regensburg.de/ (accessed on 9 January 2018).

24. COUNTER. Available online: https://www.projectcounter.org/ (accessed on 30 January 2018).

25. Zeitschriftendatenbank. Available online: http://zdb-katalog.de/index.xhtml (accessed on 9 January 2018).

26. Highwire. Available online: https://www.highwirepress.com/ (accessed on 9 January 2018).

27. Directory of Open Access Journals. Available online: https://doaj.org/https://doaj.org/ (accessed on 9 January 2018).

28. Forschungszentrum Jülich, Zentralbibliothek, Open-Access-Barometer. Available online: http://www.fz-juelich.de/zb/EN/Expertise/open_access/oa_barometer/oa_barometer_node.html (accessed on 9 January 2018).

29. Fraunhofer Gesellschaft. Allianz der Wissenschaftsorganisationen. Available online: https://www.fraunhofer.de/de/ueber-fraunhofer/wissenschaftspolitik/StellungnahmenAllianzderWissenschaftsorganisationen.html (accessed on 30 Janualy 2018).

30. OA 2020 DE. Nationaler Open Access Kontaktpunkt. Über. Available online: http://www.oa2020-de.org (accessed on 30 January 2018).

31. Mittermaier, B. Datenarbeit und "Nationaler Kontaktpunkt Open Access"—Ein Interview mit Dr. Bernhard Mittermaier. *ABI Technik* **2017**, *37*, 293–296.

32. German Federal Ministry of Education and Research (BMBF). Open Access in Germany. The Strategy of the German Federal Ministry of Education and Research. Available online: https://www.bildung-forschung.digital/files/170419_Open_Access_english.pdf (accessed on 11 February 2018).

33. GOKB—Global Open Knowledge Base. Available online: http://gokb.org (accessed on 30 January 2018).

34. VIAF—Virtual International Authority File. Available online: https://viaf.org/ (accessed on 30 January 2018).

35. Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz). Projekte: LAS:eR. Available online: https://www.hbz-nrw.de/projekte/LASeR (accessed on 30 January 2018).

36. Nationaler Statistikserver. Available online: https://statistik.hebis.de/stats/site/login (accessed on 30 January 2018).

37. Kompetenzzentrum Bibliometrie. Dateninfrastruktur. Available online: http://www.forschungsinfo.de/Bibliometrie/index.php?id=infrastruktur (accessed on 30 January 2018).

38. Universitätsbibliothek Bielefeld. BASE Bielefeld Academic Search Engine. Available online: https://www.base-search.net/ (accessed on 30 January 2018).

39. Open Citations. Available online: http://opencitations.net/ (accessed on 30 January 2018).

40. FIZ Karlsruhe. Nationales Hosting Elektronischer Ressourcen (NatHosting). Available online: https://www.nathosting.de/download/attachments/2490483/NatHosting_Projektbericht_mit_Anlagen.pdf?version=1&modificationDate=1476708609000&api=v2 (accessed on 5 January 2018).
41. NISO. Standardized Usage Statistics Harvesting Initiative (SUSHI) Protocol (ANSI/NISO Z39.93-2014). Available online: http://www.niso.org/standards-committees/sushi (accessed on 5 February 2018).
42. Clarivate Analytics. Available online: https://clarivate.com/products/web-of-science (accessed on 30 January 2018).
43. Geschuhn, K.; Stone, G. It's the workflows, stupid. What is required to make, 'offsetting' work for the open access transition. *Insights* **2017**, *3*, 103–114.
44. Mittermaier, B. From the DEAL eEngine Room—An Interview with Bernhard Mittermaier 2017. Available online: http://libreas.eu/ausgabe32/mittermaier_en/ (accessed on 30 January 2018).