

Using the institutional repository to store data related to payments

Gernot Deinzer

<https://orcid.org/0000-0002-7462-3847>

University of Regensburg

Abstract

At the University of Regensburg, all payments for publications must be made by the library. The fees, invoices and additional payment information are stored together with the document in the institutional repository. We present the motivation for this decision, the related workflow, and benefits from storing data in the repository.

Introduction

Open access has become a successful business model publishing scholarly information and is adapted by nearly all publishing houses around the world. By increasing the number of open access publications one immediately asks about the prices keeping in mind the price increase of serials. At least one should know about the spendings for open access and publishing in general. Even this is not a very easy task, because publishing fees are in the normal case not addressed to the library or the university, but to the corresponding author himself.

Preliminary considerations

Looking at the requirements from funders¹, politics², executive boards of universities and other related stakeholders, universities have to deliver a huge amount of numbers related to open access. These cannot easily be evaluated due to the fact, that not all numbers are stored in the same place and searches over platforms must be performed with the risk of losing some data.

Typical questions regarding open access are

- How many publications are open access?
- Why are publications open access?
- How is the quality of open access publications assured?
- What is the total spending for open access publications?
- What is the amount paid by central funding from the library?
- Can data be reused by publishing with suitable licenses?

These questions can be asked on an institutional level as well as over different institutions e. g. federal states. To answer these questions, one must either collect the data on the level of the whole institutions and to provide these data in a standardized way to allow third parties to collect and gather the respective information.

Open access situation

¹ E. g. Monitoring des DFG-Förderprogramms "Open-Access-Publikationskosten" – Datenschema – Erläuterungen, 2023, <https://www.fz-juelich.de/de/zb/downloads/open-science/monitoring-oa-publikationskosten-datenschema-erlaeuterungen-2023-01-25.pdf/@@download/file>

² See e.g. Kindling, Maxi, Neufend, Maïke, Stiller, Juliane, Trkulja, Violeta, Kobialka, Sophie, & Wenninger, Agnieszka. (2021, September 29). Open-Access-Landesstrategien und der Weg zur offenen Wissenschaft – Möglichkeiten und Herausforderungen am Beispiel Berlin. Zenodo. <https://doi.org/10.5281/zenodo.5535754>

There are a lot of possible reasons why a publication is open access and how we are spending resources for it. This can be either be money or staff cost. Mainly we distinguish four possible ways:

Primary open access publications at publisher site:

This involves a payment for each publication published open access. These can be payment for article in open access journals³ or buying a free license to an article in a closed journal⁴. Books can often be published openly by paying a so-called book processing charge. But even in contracts with publisher payments for each article can be included. These are the so-called Transformative agreements. Best examples are the German DEAL⁵ contracts, where each article is assigned with Publish and Read Fees. Transformative agreements with a Read and Publish model⁶ we would classify herein, because even if the contract is based on a subscription fee, all or a certain number of open access publications are included in the total fee.

Support of community-based business models

We pay a specific amount for a certain period and all publications within this are open access. OpenLibrary of Humanities (OLH)⁷, SciPost⁸, SCOAP³⁹, PLOS Community Action Publishing¹⁰ are well known examples for such a business model. But also, pledging models like KOALA¹¹, supporting of infrastructures like the DOAJ¹² we would allocate in this category.

Parallel Publication

Further on we consult our researchers making their publications open access available in the institutional (or a subject based) repository. Even if we don't spend money directly for the publication, we invest in library staff. They consult the researchers by answering legal questions and support assuring publishing rights in agreements with publishers.

Institutional offers

Beside this we provide an infrastructure to publish at the university. Within this infrastructure we offer a service as a publisher, where the primary publication is both open access and printed. For publishing journals open access we run a platform at the university. Theses, research data, and software can be published in the institutional repository. We don't charge the authors or editors any fee, even if we invest in staff and the technical infrastructure, e. g. servers and storage.

Information budget

The information budget¹³ is defined as the whole spendings for information material such as licenses to journals and eBooks, buying printed copies of journals or books, paying fees for publishing, providing inter library loan or pay per view. Just at the first glance one sees immediately

³ Gold Open Access journals are journals where the complete content is published open Access.

⁴ Journals with the choice of closed and open publishing are known as hybrid journals.

⁵ See <https://deal-konsortium.de/en/about-deal>

⁶ Read and Publishing business model means, that the payment is based on the subscription fee and a possible additional fee for publications with an open license.

⁷ See <https://www.openlibhums.org/>

⁸ See <https://scipost.org/>

⁹ See <https://scoap3.org/>

¹⁰ <https://plos.org/resources/community-action-publishing/>

¹¹ See <https://projects.tib.eu/koala/en/>

¹² See <https://doaj.org/>

¹³ See Wissenschaftsrat (2022): Empfehlungen zur Transformation des wissenschaftlichen Publizierens zu Open Access; Köln. DOI: <https://doi.org/10.57674/fyrc-vb61> and Mittermaier, B. (2022). Das Informationsbudget: Konzept und Werkstattbericht. O-Bib. Das Offene Bibliotheksjournal / Herausgeber VDB, 9(4), 1–17. <https://doi.org/10.5282/o-bib/5864>

that it is very hard to identify all payments regarding this budget. A lot of payments are directly done by authors either via the university administration or even worse by themselves. We focus ourselves here on the payments directly for publishing. All other payments (mainly for reading) are much easier to determine. We assume that all the acquisition and licensing of information material is done by the library and so all the numbers are well known. Therefore we have a closer look to all the different parts related to publishing. First, we have the classical payments for a single article. This can be an APC for publishing the article open access either in a gold open access journal or a hybrid journal, charges applied for different reasons like color images (often referred as color charges), due to the length of the article (page charges) or even additional services like processing it to the peer review (submitting charges), printing it on the cover (cover charges) or ordering author copies. There is no difference for articles in conference proceedings. Even for books there exists nowadays, besides the classical author charges for printing a copy, new charges for publishing open access. These payments are hardly to identify in the accounting system of a university. In an ideal world all invoices for information would be covered by the library and thus easily identifiable. But the real world differs. Most libraries are not paying for all information especially for the charges related to publishing. There are e. g. charges which are not covered by the library because they do not fulfill the funding requirements. Often the hybrid APCs are excluded due to the so-called double dipping issue. Although payments not related to open access are not fundable. Often open access fees like APCs are just known to certain range, because requirements are here too strict to apply like a certain maximal threshold. Sometimes authors just don't apply for a funding. This might be the case where funding grants still must be spent. In real life we have therefore a situation that charges related to publications are paid from different sites: publication fees are paid either by the library, by the researcher from their own budget or by funding bodies like the Deutsche Forschungsgemeinschaft (DFG).

Central Invoicing

To get an overall insight, how much was paid for publications at the university, we had a look at the process of buying printed books and compared it with the workflow for paying publication fees. In both cases there is the possibility of saving money due to certain agreements or contracts with publisher or book sellers. The workload will be decreased due to gathering invoices on a regular base. And often these are the same contract partners, the publisher. The question was: why is there a centralized process for buying books and why does no equivalent workflow exist for publication fees?

The simple answer to this question is the inventory of the printed book. Each book gets a stamp that it is owned by the university. In an electronic publication this is not possible and thus the publication has not to pass the central office with the stamp, in particular the library.

At the beginning of the year 2021 a change was introduced. The process of paying publication fees followed the workflow of buying books. The executive board of the university decided that all invoices regarding payment for publication must pass the library. This was done due to the recommendations¹⁴ of both the German Rectors' conference (HRK) and the library committee of the university. At the same time the administration of the university introduced the central billing invoice for all invoices of the university. This includes the publication fees. A new process for paying publication fees was installed. The invoices must be sent to the central billing office directly by the publishers and are then assigned to the library. This is done either by stating the library in the billing address or automatically if there are certain key words in the invoice (like APC, page charges or color charges) or if the billing partner is a known publisher. Afterwards the library gets in contact with the author and clarifies how to pay the invoice. This can be done from the central publication fund, from projects founded by funders like the DFG or by the resources of the

14Rundschreiben Nr. 018/2020 ("Fortführung es DEAL-Vertrags mit Springer Nature") vom 30.06.2020

researchers. Even a splitting of the invoice over different resources is possible. The library pays the invoice and makes all the internal money transfer. A copy of the publication will be stored in the institutional repository. Open access if the publication was published with a reusable right or with restricted access if the publication is closed access or has no reusable rights assigned. This corresponds to the inventory of a book.

Metadata of publication

If we are looking at the metadata of a publication, we can divide it in several distinct parts¹⁵:

- Bibliographic part: all the data describing the publication like author, title, keywords, abstract
- Technical Metadata: technical information about the publication like filetype, filesize, checksums but also the dates (submitting, accepting, publication)
- Legal metadata: this are mainly the rights assigned to the publication like the licenses
- Organizational metadata: this includes all the relationship of a publication with an organization like a research institute, the groups therein, projects, and funders.

These are the well-known metadata, and they are recommended to be stored for each publication in repositories. But if we are broadening our view on publications and with all the new publishing models, we can enlarge the set of metadata with information about payments. These can include payments directly for the publishing like APCs, page charges etc. or information about the publication being part of a contract which allows the article to be published open access or to give the article distinct reuse options. Best known examples for such contract are the so-called transformative agreements like DEAL, memberships like SCOAP³ or consortia models like OLH or SciPost.

Adding metadata to a publication

For every publication it must be clear, whether it is open access or not and, if it is open access, why it is open access.¹⁶ Therefore, first of all we specify why an article is open access or if there are no open access possibilities for it. This can be either the information that it was paid on a single article level (APC in gold or hybrid journal with distinction paid by the University of Regensburg or another research institute), no payment due to the business model of the journal (gold OA journals without APCs, diamond OA), part of a membership or transformative agreement or parallel publication due to some transferred rights (German copyright law, policies of publisher). Further on, if a payment was done by the university, the details of the payment are also added to the metadata. These are the total costs, costs related to open access, the part paid by central funding as well as the date of payment and the internal transaction numbers for each payment. The information of the invoice is added to the metadata. This consists of the invoice date, the amount and currency printed on the invoice, the invoice number and if possible, the article reference number. We also specify if there is an acknowledgement to a funder or the reference to research data as these are sometimes prerequisites for a funding. With this information we can connect the open access and payment information of a publication to the publication itself from authors from the university.

Gathering information

15 See Pampel, Heinz (2019): Auf dem Weg zum Informationsbudget: zur Notwendigkeit von Monitoringverfahren für wissenschaftliche Publikationen und deren Kosten; Arbeitspapier, Potsdam : Helmholtz Open Science Koordinationsbüro, 15 p.
<https://doi.org/10.2312/os.helmholtz.006>

16 See DINI AG Elektronisches Publizieren (2022): Positionspapier: Dokumentation der Rechtsgrundlage für Veröffentlichungen in Repositorienmetadaten, <http://dx.doi.org/10.18452/23191>

To get the information even outside the institutional repository, one must connect the different blocks of metadata. To establish a connection, one needs a unique and persistent identifier. For published articles often this identifier exists in form of a DOI and everything can be linked correctly. But for non-published articles or books, no such identifier exists, and the linking is extremely difficult. But as long as DOIs are crucial for a long-term citation of a digital object, each scholarly publication independently if it is an journal article, a book, a book section or a conference proceeding should have an assigned DOI. If a publication has more than one DOI, e.g. because the preprint, the publications and a parallel version in a repository have a distinct DOI, one should establish a connection of the different DOIs like “is preprint of”, “is equivalent to”, “is version of” etc. It makes in our opinion completely sense to assign DOIs to digital copies in a repository as far as the DOI points to exact one digital object and not to equivalent class of objects.

Displaying the information of open access

A big advance to have all the data (organizational, bibliographic, technical, legal information, payments) in one place (the institutional repository) is the possibility to analyze the data in depth. No limits are set to combine all the information and a broad overview over publishing activities can be generated. This includes e. g. costs and business models for different subjects, distribution of costs within the research institution. Money spent for different publishers, relationship between different open access business models and subjects and so on.

Even reports for executive boards and funders can easily and automatically generated. This is extremely useful to argue for central funding of fees related to open access as well as to comply with funding requirements, e. g. for the DFG-program “Open-Access-Publikationskosten”.

Exchanging data

Of special interest is the aggregation of data over several institutions. This has already been done for APCs in the project openAPC¹⁷. To our knowledge this is the first initiative to collect data about publication charges and gather the information in one place. To participate one must deliver the data in a special form defined by the project. The restriction to APCs leads to a very small set of data to be provided. By extending this set to all payments for publishing, which is not an easy task, new possibilities to analyze payments and money flows can be generated. But this data must be provided with an open license like CC0 and in a standardized and applicable way. Metadata is usually exchanged between different systems with special xml-schemes, which are provided over a common interface the so called OAI-Interface¹⁸. Even or payment information the way can and should be used as service provider can easily harvest the data from the different institutions (data provider) and build upon this data new services beginning with large databases ending with different statistical tools. Even the mechanism of constraining to specific sets, would allow stakeholders to get just the information they are interested in. Up to now no schema for payments analogous to well established schemes for bibliographic metadata like Dublin Core¹⁹ or Datacite²⁰ has been defined.

Summary

At the University of Regensburg we established a centralized workflow to process payments for publishing. All the payment information is stored as metadata related to a publication in the institutional repository. Even this is combined in the first step with additional work in the open

¹⁷ See <https://openapc.net/>

¹⁸ See <https://www.openarchives.org/>

¹⁹ See <https://www.dublincore.org/>

²⁰ See <https://schema.datacite.org/>

access team, the benefits in generating automated reports and statistics is enormous. By storing the data in a structured way in an open accessible platform the data is available for anyone who is interested in processing cost data. New evaluations of payments to publishers are possible if libraries worldwide provide this information in a standardized way. Stakeholders like funding bodies can get a deep insight about publication spending over different institutions.

Literature

Kindling, Maxi, Neufend, Maïke, Stiller, Juliane, Trkulja, Violeta, Kobialka, Sophie, & Wenninger, Agnieszka. (2021, September 29). Open-Access-Landesstrategien und der Weg zur offenen Wissenschaft – Möglichkeiten und Herausforderungen am Beispiel Berlin. Zenodo. <https://doi.org/10.5281/zenodo.5535754>

Pampel, Heinz (2019): Auf dem Weg zum Informationsbudget: zur Notwendigkeit von Monitoringverfahren für wissenschaftliche Publikationen und deren Kosten; Arbeitspapier, Potsdam : Helmholtz Open Science Koordinationsbüro, 15 p. <https://doi.org/10.2312/os.helmholtz.006>

DINI AG Elektronisches Publizieren (2022): Positionspapier: Dokumentation der Rechtsgrundlage für Veröffentlichungen in Repositorienmetadaten, <http://dx.doi.org/10.18452/23191>

Wissenschaftsrat (2022): Empfehlungen zur Transformation des wissenschaftlichen Publizierens zu Open Access; Köln. DOI: <https://doi.org/10.57674/fyrc-vb61> and Mittermaier, B. (2022). Das Informationsbudget: Konzept und Werkstattbericht. O-Bib. Das Offene Bibliotheksjournal / Herausgeber VDB, 9(4), 1–17. <https://doi.org/10.5282/o-bib/5864>

