

New metrics & open access transition

Introduction

Open access transition agreements

The National Library of Luxembourg's consortial department¹ has spent a couple of years to rollout an integrated, data-driven administrative structure that allows the design of new models and flexible shifting of costs. In 2019, we build infrastructure for producing and analysing new usage statistics on article level, overcoming inherent shortcomings of COUNTER statistics². The goal is to develop agreements with key publishers that integrate open access costs ("Publish") with current subscription agreements ("Read").

Two guiding principles:

1. Our subscriptions contain more and more Open Access content, hence the subscription costs should fall proportionally, the "Transition credit";
2. Our consortial partners pay increasing amounts for Open Access publications, these costs are covered by the savings of the subscription part.

Goals for transition agreements:

1. Transparent and sustainable for both publishers and libraries
2. Long term commitment (3-5 years)
3. Data-driven with new data on actual usage of publisher content

We aim to communicate our results pro-actively to the most relevant publishers, consortial partners, ministries and directly to scientists³. Therefore, we must use meaningful data points that are not bound by confidentiality.

The "new metrics" project is also an exploration and infrastructural base for future services, out of scope for this phase. Insights on the value of reading and publishing are available for all publishers, not just those that are the targets for open access transition agreements.

¹ Consortium Luxembourg is serving a number of consortial groupings and individual participants, including academic, research, national, public and governmental libraries, through a centralized infrastructure. The total yearly licence cost is ca. 2,5 mio EUR. Staff is 3.2 FTE. Detailed information: www.consortium.lu/about

² "Looking under the COUNTER for overcounted downloads", <https://escholarship.org/uc/item/0vf2k2p0>; "Effects of Publisher Interface and Google Scholar on HTML and PDF Clicks: Investigating Paths That Inflate Usage", <https://doi.org/10.1016/j.acalib.2018.09.014>

Reading/Subscription costs

We do not know how much Open Access content our subscriptions contain and we do not know which individual articles are read. This data is not provided by publishers. After a successful pilot project, we developed a new metric⁴ as an indicator of subscription value. The reference, historic, model is based on publisher-provided COUNTER usage data.

Example (with fake numbers)

COUNTER usage statistics for a package of ejournals, year 2018

50.000 articles (JR1) were downloaded, cost per article 2 EUR

News metrics analysis of actual number of articles used

	Total number of articles ⁵	50 400	100%
a	Only 2018 publication year	39 452	78%
b	Only non-OA content	33 451	66%
c	Only one download per user	25 245	50%

Details for the new metrics analysis:

- First we limit to a given publication year (similar to COUNTER JR5 reports). Most subscription agreements only concern the current year's new publications, past years are offered permanently for free. Past years are therefore excluded for proper assessment of the yearly subscription cost.
- Open Access articles are excluded, based on querying the Unpaywall API⁶. Only version of record articles (Gold or Hybrid OA) are excluded for now. (It would be trivial to add, for informational purposes, available Open Access pre- or post-prints of the publisher version.)
- COUNTER rejects repeat downloads within 30 seconds⁷. Per design, most publisher article landing pages require an HTML download before the PDF link becomes available. We have therefore extended the repeat download time frame to the duration of a user session, irrespective of the article format (PDF, HTML).

The "new metrics" analysis shrinks the number of downloads to half the COUNTER number. Cost per article therefore doubles. Based on this information, the cost of the subscription should be half its original price. This substantial "Transition credit" is to be used for discounts or paying publishing costs (see below). The above percentages and the amount of the "Transition credit" are all non-confidential and can be used in communications to stakeholders. Numbers can be targeted for specific subject areas⁸ to make it easier to address scientists directly.

⁴ The new metric is based on analysis of usage data at article level, using ezProxy log-files as source data, article identification by ezParse (www.ezparse.org), enrichment by Crossref and Unpaywall and final analysis by Kibana.

⁵ COUNTER and local Proxy ways of counting total number of downloads will not exactly add up

⁶ Also useful could be Crossref, DOAJ, WoS, Scopus and Romeo/Sherpa APIs, which are currently out of scope.

⁷ COUNTER Code of Practice, Release 4, Appendix D, Section 5 <https://www.projectcounter.org/wp-content/uploads/2016/02/APPD.pdf>

⁸ Subjects of the used articles are available through metadata enrichment. With some manual curation, this allows slicing of the data not only by consortial member and publisher but also by scientific subject.

Publishing costs

Knowing the total cost

It is notoriously difficult to identify all payments made to publishers outside of subscription costs, although publishers themselves are increasingly able and willing to help. These include Open Access publishing fees (APC⁹) but also page and colour charges (irrespective of Open Access status). We incentivise CFOs of consortial members to provide such costs, as without knowledge of costs, we cannot recoup them. It is important to record these expenses following their budget years and not rely on publishing dates only. Establishing these workflows is the most difficult and time-consuming but largely pays for itself if the recoupment model is successful.

Knowing local publications

Local research output easier to obtain, the only major issue is with pre- and post print copies harvested from repositories without proper identifiers, such as DOI. Versioning of article copies is an ongoing problem, as is quality control for metadata (aka cataloguing...).

Calculating and spending the Transition credit

The “Transition credit” identified above (in this example roughly half the current subscription cost) is calculated ex-post, for a calendar year or an arbitrary 12 month period ending in time for the renewal discussions¹⁰. To calculate the corresponding publishing spent, all expenditures are simply pro-rated to this period.

As all analysed periods and amounts are in the past, we can negotiate with facts and adjust payments year after year. Minimum & maximum variation ceilings can be defined to facilitate long-term planning security for both sides.

The transition credit is entirely used to pay for any past publishing costs (APCs plus colour/page charges). Remaining credit (or due additional spent) is negotiated. If no discount is given, the publishing costs are “virtually” increased and may suddenly look rather expensive. As this is a “virtual” spent, it is also not confidential (nor should publishing costs be at all). Again, these numbers can be aligned to subjects and consortial members for targeted communication towards CFOs and scientists.

If a remaining credit is applied to publishing costs, they would apply to future expenses, as past publishing costs have already been paid. The model can adapt to such realities by lowering the price charged by the publisher for next years publishing costs or adopt a model of pre-payment or pre-negotiated prices. It is essential that such negotiated publishing costs are not confidential. Publishers may want to push in the direction of pre-negotiated “big-deal” publishing cost models, but that way we miss the other publishing costs (page/colour charges) and risk paying for nothing if the number of articles is not reached. As Luxembourg has a comparatively tiny research sector, such deals may be less appealing to us.

⁹ APC = Article Processing Charge. Used for the cost to publish in Gold and Hybrid Journals.

¹⁰ To have numbers ready for the 2019 renewal discussions, we would take the twelve month period from 1st September 2017 until 31st August 2018. This is the same period we use to apply a division key to split costs between members of some consortial groupings. This division key is based on traffic in MegaBytes per product per member for a 12 month period and applied to next year’s cost of each subscription.



Further data services

As mentioned in the beginning, the “new metrics” project is also an exploration and infrastructural base for future services, currently out of scope.

Having article-level usage data, licence lifecycle information and local publications in one system allows for relevant insights based on citation analysis of both sets of articles, for the consortium, research managers and scientists.

Example: Which titles are cited most by local scientists? Are titles cited which are not subscribed?

Incoming citations add another level of insight: Do we subscribe to all titles, which cite local publications, are we missing any? Further citation analysis to identify of so-called “hidden institutes” and possibly “citation cartels” are further possibilities.

Producing and sharing of such data internationally would increase their value immensely and contribute to the discussion around data owner- and stewardship in future research infrastructures.

Identifying how many unique articles are downloaded per year, and related to this, how often each article is downloaded, points to an interesting, if longshot, evolution of subscription licences towards DDA/PDA¹¹ like licences for articles, similar to ebook purchases.

Requirements for consortial management system

We analyse reading and publishing costs of the past year and negotiate discounts based on that analysis for future, upcoming costs. This requires centralised and flexible budget allocation mechanisms between consortial partners, for example through the use of an invoicing intermediary (either internal or external). A management tool¹² that can handle the subscription lifecycle, importing and reconcile with subscriptions the local publications and usage is necessary. Moving into publishing costs, a new cost type is required as publishing costs are not necessarily related to any existing publisher licences. During 2019 we will define the scope of which data are best managed in which systems. ConsortiaManager.com is our partner for this task.

Interest of publishers

Detailed usage data of their competitors’ content may be of interest and value for publishers, especially if larger library networks start building up such datasets.

Contact

Patrick.Peiffer@bnl.etat.lu
www.consortium.lu/about

¹¹ PDA=Patron Driven Acquisition; DDA=Demand Driven Acquisition. A purchasing model where a large catalogue of ebooks is offered and purchases are triggered by actual usage by library users.

¹² In Luxembourg we currently use ConsortiaManager.com, together with Primo, SFX and Leanlibrary.com as access tools.