

Report for Jisc Collections on Total Cost of Ownership Project:

Data Capture and Process

By

Dr Hazel M Woodward
hazel@informationpower.co.uk

Helen L Henderson
helen@informationpower.co.uk

*Information Power Ltd
5 Stirling Acres Road
Kirkcudbright DG6 4EY*

www.informationpower.co.uk

May 2014

Contents

1. Background 1

2. Consolidation of data within HEIs..... 1

3. Problems with files received..... 1

 3.1 Excel spreadsheets..... 1

 3.2 Pricing/ Costs 1

 3.3 Licensing..... 2

 3.4 Data elements..... 2

4. Data elements/metadata essential and desirable for monitoring APC expenditure 3

5. Data sources for information about publications..... 4

6. Terminology/definitions needed 4

7. One-off capital costs and ongoing costs 5

8. Problems with Jisc APC data system..... 5

9. Recommendations for advocacy tactics within HEIs 5

1. Background

Information Power Ltd (IPL) were contracted to undertake a data collection exercise during March and April 2014 in order to obtain real data for Jisc Collections to input into their modelling tool for publisher negotiations for Total Cost of Ownership (TCoO). During the data collection exercise a number of issues arose which were felt significant enough for Jisc and Jisc Collections to ask for further feedback on the data collection and the processes. It must be remembered that the briefing for this specific project was data collection for the sole purpose of Jisc Collections negotiations, but the comments on the report extend beyond that specific data requirement. This report gives brief details of the main issues that arose regarding both the data collection and comments from HEIs participating in the project.

2. Consolidation of data within HEIs

One of the most significant issues which arose during the TCoO data capture exercise was the fact that not one of the 24 universities involved in the project were able to tell the project the *total* number of APCs paid by their institution. In the vast majority of cases, the Library handled the RCUK and Wellcome funds and captured data on APCs paid from those funds. But none were able to establish how many APCs were paid from departmental/ faculty research project funds. A number of librarians asked the University Finance department for this information, but none were able to provide a figure. The speed that the data was collected at mean that institutions had very little time to compile data which they were not already collecting, which also meant that in some cases the data could not be fully checked for accuracy.

In addition it was interesting to note that, despite a specific request, only one of the libraries was able to provide IPL with the total number of articles published by their institution in a given academic year.

3. Problems with files received

3.1 Excel spreadsheets

Many libraries were clearly inexperienced with working with Excel spreadsheets. There were frequently workbooks with multiple tabs each with different column layouts and data

Dates and costs were entered differently, even in the same sheets, with both entered as text, or with non-numerical characters. Notes and comments were frequently included in fields, meaning it was sometimes difficult to decide if issues were resolved.

3.2 Pricing/ Costs

One of the main issues was VAT. It was not clear in most cases whether VAT was included in the APC or subscription charge. When asked, many institutions said "Finance deal with all that".

Costs were frequently stated as estimates, and although there was a sometimes a column for actual cost or invoiced cost, this was mostly not filled in. Foreign payments were sometimes given in local

currency, sometimes converted, and, again the VAT was not at all clear. Many records had no final GBP figure. Where no costs were given, the figure was added based on Finch estimate for APC cost of £1,470.

There were occasional problems with extra payments demanded for producing invoices by some publishers (e.g. IEEE and the National Academy of Science charged £50 for an invoice) and it is not clear if these were included or not in the final APC cost.

Sometimes publication was blocked because the publisher claimed money had not been received. Most publishers were not prepared to publish until all due costs had been received. Moreover, it seems that some university accounts departments can hold payments up, thus delaying publication.

Pre-payments for APCs caused a lot of confusion. Where pre-payment had been made, the APC cost was not always shown in the spreadsheet. Because some of these have different rates according to number of articles the effective cost can be calculated and should be shown.

Membership and discounts: for some society publishers there is an APC member's discount. Some HEIs pay the membership fee to get the discount. There are other discounts for quantity of articles and ties with subscription packages. These cause a big variation between HEIs for the cost of APCs for a given title.

The data collected by most libraries would not be adequate for any auditor, as there was no clear trail to final publication or evidence of the publisher meeting funder requirements.

3.3 Licensing

There was a lot of confusion by both publishers and libraries over the difference between the various business models and licensing models. Frequently it was unclear whether the APC payment was made for an article to be totally open, embargoed, available for deposit in a repository, etc. There were also many references to the publisher not meeting OA copyright requirements, with refunds or offers of errata slips being made by the publisher.

It is exceptionally confusing to try and calculate any standard APC charges because every publisher has different terms and condition (e.g. number of pages, number of words, additional charges e.g. for images). The only central source of this data that we have been able to identify is SHERPA/ROMEO but this is only partial data and is not kept up to date.

3.4 Data elements

For the TCoO Project IPL asked for: the name of the institution; publisher and title of journal; the publication date; the total cost of the APC; and the funder. We did not manage to get these for all the records we were given, in part because some articles were awaiting publisher or funder approval, what follows is a list of the issues which arose with the data received from the volunteer libraries.

1. Some APC records turned out to be for books, book chapters, and conference proceedings. There should probably be a separate category for these, as they will be on the increase and included in future funding.
2. There was frequently no indication of the journal to which the article had been submitted.

3. Where there was article level data (title or author) some of the gaps were able to be filled by IPL. However, this was very time consuming.
4. The funder was frequently not shown in the data. For the purposes of the project, where the funder was not specified the assumption was made that it was RCUK.
5. Some APC spreadsheets indicated whether the title was Green or Gold but this was often incorrect.
6. Publication date was not always given - just the DOI/PubMed ID. IPL had to use the article title and authors to get a publication date. Where the date *was* given, it was not clear if this was for first publication (usually electronic) or print. Normally when we looked up publication date we used the electronic. Note: This will only be a problem if the difference between calendar and financial years is an issue.
7. There were comments in the APC data saying “no time to check if article had been published”. Since they often had the DOI or PubMed link this should have been easy to do. When there was no DOI but there was an article title, it was straightforward to use Google to get the DOI and /or PubMed links and publication date off the publishers’ web site. This provided an opportunity to see how the publisher was presenting the OA article and the relevant licence terms. They were often not at all clear!

4. Data elements/metadata essential and desirable for monitoring APC expenditure

IPL believe that Jisc Collections should think here about the wider need for data on HEI APC expenditure data. It will certainly be required by university accounts departments, auditors, Research Gateway, funders, the authors, university departments and the university library.

The following is a list of what IPL believe should be collected and documented in a consistent way:

- Date of initial application by author
- Journal Title
- Publisher (to check for transferred titles, or subsidiary publishers e.g. Cell Press)
- ISSN (with clear definition of whether it is the p-ISSN, e-ISSN or i-ISSN)
- Article Title
- Category of article: e.g. journal, conference proceeding, book or book chapter.
- Corresponding Author
- Other authors (where appropriate, given that large numbers of these will cause processing issues)
- Sharing APC cost with another institution
- University Department (or something to show which part of the HEI the article is coming from)
- Funders name
- Research project ID
- Status with Funders (in some cases publication is on hold while applying for permission to publish by the funding agency)

- Internal HEI Accounts ID or cost code (this frequently allows payments to be tracked and consolidated)
- Financial data elements: estimated cost, cost in actual currency (e.g.\$ or euro), conversion rates and final GBP figure, VAT, and any additional costs such as cost of invoice, additional page charges, costs for images or tables
- Final APC payment in £s with VAT
- Embargo period (where applicable)
- Publication date: month and calendar year
- Licence terms expected and checked
- Has publisher incorporated correct OA permissions statement?

5. Data sources for information about publications

IPL had quite a few problems with identifying the title or publishers as the publisher given was frequently an imprint or subsidiary publisher (e.g. Cell Press, Routledge, Frontiers). Some “publishers” turned out to be agents (such as Dartmouth Journals or Sheridan Press) or aggregators e.g. GeoWorld.

DOAJ did not help decide if a title was hybrid and it is not up to date with mainstream publishers OA offerings. The best route was checking individual publishers’ web sites for lists of fully open access titles.

As the publishers’ web sites were checked it also provided an opportunity to see if the spreadsheets’ APC costs for an article corresponded to the publishers’ statements on this. It was often difficult to decide, but it was possible to estimate some of the gaps in the APC cost data this way.

IPL is not sure when the library gets the DOI before publication - but they clearly get it in advance. There are several sources for such data which could be used to help the library’s APC staff keep tabs on when APC articles are published and if they have the correct licence terms. These include Cottage Labs Open Access Gauge; University of Nottingham’s SHERPA/ROMEO; Heriot-Watt’s JournalTOC’s database which now has article DOIs and licence details; and ZETOC from the British Library /MIMAS.

As an experiment we looked at the RCUK project *Research Gateway* - which contains data on published outcomes - for one institution. It turns out that is difficult to access information about articles as only the RCUK grant project IDs are linked to the institution. Each project then needs to be examined for the articles produced for that project. In addition there is no differentiation between the lead institution and collaborating institutions. As a result of the investigation of this institution 200 published articles were identified, of which only 1 appeared in the APC data.

6. Terminology/definitions needed

IPL believe that there are 3 important areas where terminology needs urgent clarification. These are:

- Confusion between Green, Gold and hybrid. A lack of understanding between licensing models and business models

- Absolute clarity about what is paid to the publisher on what terms
- The publishers exact terms of the licence.

7. One-off capital costs and ongoing costs

For Phase 3 of the TCoO Project, libraries were asked to estimate their initial set up costs for administering APC payments and their ongoing costs. It was clear that no library had actually collected this data, and all respondents said their answers were very much an estimate. Of the 13 libraries who responded, 4 said they had no initial set up costs. Of the others, the set up costs ranged from a modest £208 (for staff training) to £24,678.

Ongoing costs were identified as: consulting/advising authors; communication/fact finding with publishers; financial management/administration of payments; internal meetings/ administrative planning; dissemination of information on APCs and OA; and liaison with funding bodies.

Each institution identified their 3 busiest months and calculated staff costs which ranged from £797 to £23,915.

8. Problems with Jisc APC data system

IPL understand that Jisc Collections are aware of the shortcomings of the OAK system. One problem area reported to IPL was, according to one HEI, OAK does not deal with foreign VAT payments (reverse charge issue) leaving it to the HEI to sort out, but this makes it difficult because the records returned don't have the article title (which is the only way to identify some articles pre-publication). Is there a need for an identifier for articles prior to publication?

9. Recommendations for advocacy tactics within HEIs

Train all staff working on the various tasks associated with APC payments staff. This is a significant opportunity to strengthen liaison with individual researchers, faculty and other administrative departments (e.g. Finance & Research Office) and to promote OA.

In a breakout at UKSG Manchester University showed how this could be done by creating a cross-departmental team including the Research Office, Finance, HR and Library all working together to create a bespoke integrated system which allowed and encouraged the authors to input data by selecting from data elements provided by each department.