

Article

Open Access Article Processing Charges (OA APC) Longitudinal Study 2015 Preliminary Dataset

Heather Morrison *, Guinsly Mondésir, Jihane Salhab, César Villamizar, Alexis Calvé-Genest and Lisa Desautels

School of Information Studies, University of Ottawa, 111-08 Desmarais Bldg., Ottawa, ON K1N 6N5, Canada; gmond071@uottawa.ca (G.M.); jderg011@uottawa.ca (J.S.); cvill022@uottawa.ca (C.V.); acalv018@uottawa.ca (A.C.-G.); LDESA094@uottawa.ca (L.D.)

* Correspondence: Heather.Morrison@uottawa.ca; Tel.: +1-613-562-5800

Academic Editor: Sabina Leonelli

Received: 15 February 2016; Accepted: 22 March 2016; Published: 14 April 2016

Abstract: This article documents Open access article processing charges (OA APC) longitudinal study 2015 preliminary dataset available for download from the OA APC dataverse [1]. This dataset was gathered as part of *Sustaining the Knowledge Commons* (SKC), a research program funded by Canada's Social Sciences and Humanities Research Council. The overall goal of SKC is to advance our collective knowledge about how to transition scholarly publishing from a system dependent on subscriptions and purchase to one that is fully open access. The OA APC preliminary data 2015 Version 12 dataset was developed as one of the lines of research of SKC, a longitudinal study of the minority (about a third) of the fully open access journals that use this business model. The original idea was to gather data during an annual two-week census period. The volume of data and growth in this area makes this an impractical goal. For this reason, we are posting this preliminary dataset in case it might be helpful to others working in this area. Future data gathering and analyses will be conducted on an ongoing basis. We encourage others to share their data as well. In order to merge datasets, note that the two most critical elements for matching data and merging datasets are the journal title and ISSN.

Data Set: <http://dataverse.scholarsportal.info/dataverse/oaapc>

Data Set License: There is no license for the dataset as a whole, as individual elements are derived from different sources, which may have their own terms.

Keywords: open access; scholarly communication; article processing charges; publication fees

1. Summary

This dataset includes information on open access journals derived from the Directory of Open Access Journals (DOAJ), developed as the base for a longitudinal study on the open access article processing charges (APC) method used by about a third of open access journals. In the APC business model, a payment is made, by an author, institution, or funding agency, for publishing an article so that the article can be freely available to everyone (open access). In addition to DOAJ metadata, this dataset includes 2010 APC data provided by Solomon and Björk [2], a smaller set of pilot project data collected by the research team in 2013, and a fuller set of data collected on APCs by the research team in 2014 and 2015, as well as additional data relating to APC sub-model (e.g., variations in pricing, page *versus* article charges), analysis of publisher type, problematic (but possible useful) article-level metadata screen scraped from DOAJ, and a custom subject analysis. To date, these data were used as the basis for a 2014 DOAJ APC survey [3]. This project received funding from Canada's Social Sciences and Humanities Research Council under the Insight Development Grant program for 2014–2016. In

2016, the dataset will be updated and expanded to include publishers missed in 2015, at which point data analysis and preparation for a new survey article is anticipated. At present, there is keen interest from research funders, libraries, scholars, and publishers on the economics of transition to open access. This dataset will facilitate and speed up the work of other researchers, and this document describing the data is necessary to understand and analyse the data.

2. Data Description, Method and Limitations

2.1. Major Sources for this Dataset

Major sources of data for this dataset include:

- the *Directory of Open Access Journals* (DOAJ) downloadable metadata; the base set is from May 2014, with some additional data from the 2015 dataset
- data on publisher article processing charges and related information gathered from publisher websites by the SKC team in 2015, 2014 (Morrison, Salhab, Calvé-Genest and Horava, 2015), and a 2013 pilot
- DOAJ article content data screen scraped from DOAJ (caution; these data can be quite misleading due to limitations with article-level metadata)
- Subject analysis based on DOAJ subject metadata in 2014 for selected journals
- Data on APCs gathered in 2010 by Solomon and Björk [2] (supplied by the authors). Note that Solomon and Björk use a different method of calculating APCs, so the numbers are not directly comparable; please refer to Solomon and Björk [2] for details on their methods.
- Note that this full dataset includes some working columns, which are meaningful only by means of explaining very specific calculations, which are not necessarily evident in the dataset *per se*. Details below.

2.2. Significant Limitation

- This dataset does not include new journals added to the DOAJ in 2015. A recent publisher size analysis indicates some significant changes. For example, DeGruyter, not listed in the 2014 survey, is now the third largest DOAJ publisher with over 200 titles. Elsevier is now the 7th largest DOAJ publisher. In both cases, gathering data from the publisher websites will be time-consuming as it is necessary in order to conduct individual title look-up.
- Some OA APC data for newly added journals was gathered in May 2015 but has not yet been added to this dataset. One of the reasons for gathering these data is a comparison of the DOAJ “one price listed” approach with potentially richer data on the publisher’s own website.

3. Explanation of Specific Data

3.1. A to Q: DOAJ Metadata

Columns A to Q are DOAJ metadata, with the exception of titles added in 2015 that are on the publisher’s website but not in DOAJ. Most of the DOAJ metadata are from 2014 (at the time of the first annual survey). Some 2015 data were added. See, for example, column DR “in DOAJ 2015 not 2014”—any DOAJ metadata for these titles were taken from the DOAJ 2015 dataset. Note also that our 2014 DOAJ file did not include keywords; any keyword data are from 2015. Titles that were taken from the publisher’s website that were not in DOAJ can be identified using the column DT “not in DOAJ 2014 or 2015”. Titles that only have information in columns A and B (publisher and title) are another indication that the title was on the publisher’s website but not in DOAJ (Table 1).

Table 1. Columns A to Q.

Column	Column Title and Notes/Deviations from DOAJ Metadata
A	Publisher (from DOAJ; occasional clean-up to facilitate gathering, e.g., typo correction)
B	Title (almost always directly from DOAJ; occasional small variations due to title name or publisher changes)
C	Title alternative (from DOAJ; 2015 only. These data are not used in the study. Data from 2014 were deleted.
D	Identifier. Journal URL.
E	Language.
F	ISSN
G	E-ISSN
H	Keyword DOAJ 2015
I	Start year
J	End year (blank in DOAJ; titles no longer active are removed)
K	Added on date (date added to DOAJ or record updated)
L	Subjects
M	Country
N	Publication fee. Blank in DOAJ as of 2014 and 2015. Earlier DOAJ metadata indicate Yes, No, or Conditional here.
O	Further information. URL for further information on publication fee. Blank in DOAJ as of 2014 and 2015.
P	CC license
Q	Content in DOAJ. Always “Yes” in DOAJ as of 2014 and 2015. This is incorrect, as only some DOAJ journals provide article-level metadata, which this column is intended to indicate.

3.2. R to U: Publisher Size (Publisher APC Journal List) and Type (Commercial, Society, etc.)

Important limitations in regards to publisher type: In 2015, we conducted more in-depth research on publisher type than in 2014 for larger publishers. For this reason, there are more mixed publisher types. It is possible that mixed types are under-represented due to limitations in our analysis. That is, for larger commercial publishers, we assume all journals are commercial, but in some cases it takes in-depth reading about each journal to accurately identify whether a partnership is involved (Tables 2 and 3).

Table 2. Columns R to U.

Column	Column Title and Notes
Publisher size	Publisher size is the number of journals by publisher that have APCs. These data are derived either from the 2014 publisher size analysis, or the 2015 full publisher website list where additional data were taken from the publisher website. These data are used in analysis of APC journal portfolio size and to calculate the sampling factor for small APC journal publishers (less than 10 APC journals) as these are sampled on a random basis. An important limitation of these data is that they have not been updated to reflect the 2015 DOAJ metadata set.
Publisher type	Determined by members of the SKC team through analysis of publisher’s website. Codes are listed below.
Publication charges	Used in 2014 comparison of journals with confirmed publication charges and sample of 100 journals with no publication charges by publisher type.
Sampling factor	Based on publisher size. Used as a correction for calculating overall average APC to reflect sampling of smaller publishers.

Table 3. Publisher Type—Codes.

c	Commercial. Used only if commercial nature confirmed, e.g., through reading journal “about” page, notes about a registered or limited liability company on the contact page, etc.
g	Government
np	Non-profit
.com	Nothing known about publisher type, has .com in URL
.org	Nothing known about publisher type, has .org in URL
s	Society (or association)
u	University
un	Unknown
Mixed types	
c/s	Commercial/society partnership
c/u	Commercial/university partnership
un	Unknown

3.3. SKC Article Processing Charges/Article Page Processing Charges and Related Information

Table 4, below, provides a column-by-column explanation of the APC and related information contributed by the research team in 2013, 2014, and 2015.

V to AX (2014)

AY to BD (2013)

CL to DO (2015)

Important limitation: Although the list of variations (see Table 4, columns AD to DN) is long, not every variation or even every common variation is included. For example, we did not capture colour charges, which are quite common.

Table 4. Article Processing Charges/Article Page Processing Charges and related information.

Column (2014)	Column (2013)	Column (2015)	Column Title	Notes
V	AY	CL	APC	See below for details
W	AZ	CM	APC Original Currency	See APC below for details
		CN	Rate	Currency exchange rate (to 2 decimals) as of 15 May 2015
X	BA	CO	APC USD	APC in USD based on Bank of Canada exchange rate (or country's national bank where Bank of Canada data are not available) as of May 15 of the year in question.
Y	BB	CP	Max. pages per article	Maximum pages per article where indicated. This may be an absolute maximum or a maximum included under the default APC.
Z	BC	CQ	Article page processing charge	Per-page cost for journals that use this method in the original currency. See APC and APPC below for selection criteria.
AA	BD	CR	APPC original currency	See APC and APPC below
AB		CS	APPC USD	See APC and APPC below
AC		CT	APPC USD Value	Copy of column AC with formula removed
AD		CU	Variations in price (discounts, memberships) Y/N/NM (not mentioned)	Data are entered here only if there are data for APC or APPC. If ANY variation in pricing is mentioned, Y for Yes is entered. N for No means a clear-cut message that there are no variations on the publisher's website. NM for not mentioned means that there is no indication as to whether or not variations might apply.
AE		CV	Premium price for fast track	Optional additional charge to speed up publication
AF			Extra charge based on number of pages	Per-page or lump sum based on length of article. Higher per-page cost past a certain length for journals using page charges.
AG		CW	Language editing	Extra charge if copyediting needed. Sometimes optional, <i>i.e.</i> , authors can make their own arrangements or pay for this service through the journal.
AH		CX	Extra charge for repository deposit	Extra charge for depositing in an institutional or subject repository. Note that many journals do this without an extra charge.
AI		CY	Waivers/discounts based on income	Indication that a waiver may be considered in case of hardship (other than medium and low income countries)
AJ		CZ	Differential pricing for local authors	Whenever different pricing is given for authors in a particular region. In some cases "local" is assumed.
AK		DA	Waivers/discounts for low/medium income countries	A common discount, often based on World Bank country classifications.
AL		DB	Waivers/discounts based on contributions of work to journal (editing/reviewing)	"Based on contributions" is an assumption. Generally discounts refer to discounts for editors, reviewers, <i>etc.</i>

Table 4. Cont.

Column (2014)	Column (2013)	Column (2015)	Column Title	Notes
AM		DC	Waivers/discounts based on individual membership in society or association	This is for individual society members as distinct from institutional memberships.
AN		DD	Institutional Memberships	Institutional memberships is a model under which an institution such as a university pays for a membership that gives their authors either a discount or free publication.
AO		DE	Submission fee	Fee charged on submission rather than on publication
AP		DF	Discounts for manuscript/review transfer	Discount when an article has already been reviewed by another journal, e.g., if reviews are transferred and the author gives an indication of how they have addressed the reviewer's comments.
AQ		DG	Extra charge for CC-BY (or varies by license type)	Fee varies based on chosen license.
AR		DH	Temporary discounts	Special time-limited offer
AS		DI	Differential pricing by article type	Pricing varies by article type. For example, a research article may cost more than a case study.
AT		DJ	APC only if there is an author fund	APC applies only if the author has access to funding, otherwise free
AU		DK	Discounts for Students	Obvious
AV		DL	Discounts for high quality or extra charges for poor quality	Language is typically vague
AW		DM	Using publisher's template	Lower price if publisher's template is used
AX		DN	Different price format Latex/Word/PDF	Pricing is higher or lower depending on format
DN		DO	Loyalty discount	Discount for repeat publication in same journal Discount for publishing more than one article in a given journal

3.4. APC and APPC Details

Article processing charge as listed on the publisher's website, in the original currency. In the vast majority of cases, this data was gathered during the census period of 15–30 May, however, in some cases missing data were gathered in fall of 2015.

Where more than one currency is listed (this is common), we select what appears to be the primary currency, e.g., the first currency listed or the currency for local authors. Where pricing is available for different types of articles, the price for research articles is selected. Where discounted pricing is listed, the price before discounts is selected. Where a price is given for up to a certain number of pages, this is the price listed. "0" in this column indicates that a publisher clearly uses APCs for this journal but that publishing is currently free. For example, Hindawi regularly offers free publishing for their journals on a rotating basis. "No publication charge" means that we have confirmed that the journal does not have any fees associated with publication. "No cost found" means that we could not confirm whether or not there is an APC. "Title not found" means that we were not able to confirm whether the journal still exists or not.

3.5. BF to BP: PubNumber

This is a rough indication of the number of articles per journal for journals that provide article-level metadata to DOAJ. After completing the gathering of data from DOAJ, a study comparing these data with actual journal publishing data uncovered a problem with article metadata supplied to DOAJ, which makes these data highly unreliable. This is included in the full dataset on the premise that flawed data are better than no data.

3.6. BQ to BT

These are working columns for a small study comparing journals by small publishers with and without APCs as described on the Sustaining the Knowledge Commons blog:

Publication Charge (1 = yes, 0 = no)

DOAJ No Charges (Sampling)

Publisher Size (DOAJ No Charges)

DOAJ Confirmed charges (sampled).

3.7. BU to BX

Subject classification: This is an SKC grouping of subject classifications intended to roughly mirror the work of Solomon and Björk for comparison purposes.

3.8. BY to CG

Data supplied by Solomon and Björk.

3.9. CH to CJ

Working columns.

3.10. CK—*Preliminary Sample 2015—Y for all Titles for which We Had Data from 2010, 2013, or 2014, to Permit the Longitudinal Analysis*

3.11. DP to DR

Columns for entering APC data listed in DOAJ for journals, added after March 2014, that have charges. Data not yet entered.

3.12. DR to DU: *For Recording Changes in DOAJ from 2014 to 2015 Relevant to Journals Sampled*

DR: In DOAJ 2015 not 2014—Y for titles added, based on publisher website information that were not included in the 2014 sample.

DS: In DOAJ 2014 not 2015—Y for titles from our sample in 2014 that we could not find in the 2015 DOAJ metadata file.

DT: Not in DOAJ 2014 or 2015—Y for titles drawn from publisher websites that were not in DOAJ either year.

4. Using these Data (Licensing)

This dataset is derived from several sources, including the DOAJ metadata (which has its own license terms posted on the DOAJ website), other data screen-scraped from DOAJ, factual data gathered from publisher's websites, 2010 data provided by Solomon and Björk, and our team's analysis. If you are making use of our dataset as a whole, please cite: Morrison, H.; Salhab, J.; Mondésir, G.; Calvé-Genest, A.; & Villamizar, C.; Desautels, L. Open access article processing charges longitudinal study 2015 preliminary dataset [<http://dataverse.scholarsportal.info/dvn/dv/oaapc>]. If you are drawing from the other sources, please cite the other sources. There is no license for the dataset as a whole, as individual elements are derived from different sources, which may have their own terms.

Acknowledgments: The authors gratefully acknowledge funding provided by Canada's Social Sciences and Humanities Research Council under an Insight Development Grant for the Sustaining the Knowledge Commons research program of which this project forms a part.

Author Contributions: Heather Morrison: Principal Investigator, project design, supervision, primary drafter and data gatherer. Data gathering and analysis: Alexis Calvé-Genest, Lisa Desautels, Guinsly Mondésir, Heather Morrison, Jihane Salhab, and César Villamizar. César Villamizar conducted analysis using analytic

and data visualization software and standard office applications, developed the research data management plan, including the identification and classification of electronic files, naming conventions and file version control.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Morrison, H.; Salhab, J.; Mondésir, G.; Calvé-Genest, A.; Villamizar, C.; Desautels, L. Open access article processing charges longitudinal study 2015 preliminary dataset. Available online: <http://dataverse.scholarsportal.info/dvn/dv/oaapc/> (accessed on 22 March 2016).
2. Solomon, D.J.; Björk, B.C. A study of open access journals using article processing charges. *J. Am. Soc. Inf. Sci. Technol.* **2012**, *63*, 1485–1495. [[CrossRef](#)]
3. Morrison, H.; Salhab, J.; Calvé-Genest, A.; Horava, T. Open Access Article Processing Charges: DOAJ Survey May 2014. *Publications* **2015**, *3*, 1–16. [[CrossRef](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution-NonCommercial-NoDerivatives (CC-BY-NC-ND) license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).