

Data Descriptor

Open Access Article Processing Charges (OA APC) Longitudinal Study 2016 Dataset

Heather Morrison *, Widlyne Brutus, Myriam Dumais-Desrosiers, Tanoh Laurent Kakou, Katherine Laprade, Salah Merhi, Arbia Ouerghi, Jihane Salhab, Victoria Volkanova and Sara Wheatley

School of Information Studies, University of Ottawa, 111-08 Desmarais Bldg., Ottawa, ON K1N 6N5, Canada;
* Correspondence: Heather.Morrison@uottawa.ca; Tel.: +1-613-562-5800 (ext. 7634)

Academic Editor: Sabina Leonelli

Received: 8 March 2017; Accepted: 1 April 2017; Published: 8 April 2017

Abstract: This article documents Open access article processing charges (OA APC) Main 2016. This dataset was developed as part of a longitudinal study of the minority (about a third) of the fully open access journals that use the APC business model. APC data for 2016, 2015, 2014, and 2013 are primarily obtained from publishers' websites, a process that requires analytic skill as many publishers offer a diverse range of pricing options, including multiple currencies and/or differential pricing by article type, length or work involved and/or discounts for author contributions to editing or the society publisher or based on perceived ability to pay. This version of the dataset draws heavily from the work of Walt Crawford, and includes his entire 2011–2015 dataset; in particular Crawford's work has made it possible to confirm "no publication fee" status for a large number of journals. DOAJ metadata for 2016 and 2014 and a 2010 APC sample provided by Solomon and Björk are part of the dataset. Inclusion of DOAJ metadata and article counts by Crawford and Solomon and Björk provide a basis for studies of factors such as journal size, subject, or country of publication that might be worth testing for correlation with business model and/or APC size.

Data Set: <http://dx.doi.org/10.5683/SP/KC2NBV>

Data Set License: there is no specific license. See the article for details.

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

1. Summary

This article describes an update and expansion of the preliminary 2015 dataset described in *Data* [1,2]. This dataset includes information on open access journals derived from publisher websites and 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). This dataset also includes 2015 APC data provided by Crawford [3], 2010 APC data provided by Solomon and Björk [4], 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-models (e.g., variations in pricing, page versus article charges), analysis of publisher type, and a custom subject analysis. To date, these data were used as the basis for a 2014 DOAJ APC survey [5]. This project, Sustaining the Knowledge Commons (SKC) is funded by Canada's Social Sciences and Humanities Research Council. 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 analyze the data.

2. Data Description, Method and Limitations

Major Sources for This Dataset

Major sources of data for this dataset include:

- *Directory of Open Access Journals* (DOAJ) downloadable metadata; the base set is from 1 February 2016, May 2014 DOAJ metadata is included, with some additional DOAJ data from the 2015 dataset
- Data on publisher article processing charges and related information gathered from publisher websites by the SKC team in 2016, 2015, 2014, and a 2013 pilot
- Subject analysis based on DOAJ subject metadata in 2014 for selected journals
- Data on APCs gathered in 2015 by Walt Crawford [3]
- Data on APCs gathered in 2010 by Solomon and Björk [4] (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 [4] for details on their methods.

The primary purpose of this dataset is to support longitudinal study of APC pricing. For this reason, APC data from different years is presented in reverse chronological order and essential information for understanding this data is located near the left hand side of the spreadsheet while additional information for each year in question is located elsewhere in the spreadsheet. In analyzing the data it may be useful to make a copy of the spreadsheet, rearrange columns to facilitate comparison of factors under study, and possibly to eliminate additional information as the spreadsheet is large and can be slow to manipulate. Crawford's dataset is available for download as open data [3]. The following table indicates data types and sources (Table 1). The following section provides details on what is included in each column.

Table 1. Major data sources: what is included, where it comes from, and where to find it.

Data Type	Source(s)
APC and APPC (article page processing charges) data for 2016, 2015, 2014, 2013, and 2010	Primary source: publisher website lookup and analysis by SKC team. 2016 data includes data from Walt Crawford's study. Source of data for 2016 is indicated in a provenance column. 2010 data provided by Solomon and Björk. Currency conversion data: Bank of Canada 10-year currency converter and XE currency converter.
Publisher type	Publisher website analysis. Incomplete., use with caution.
Walt Crawford's 2011–2015 dataset	APC information, subjects, notes, and article counts for each journal for 2011–2015.
DOAJ metadata 2016 and 2014	Downloaded from DOAJ website and included "as is" to facilitate journal matching, e.g., if URLs are found to be incorrect in the DOAJ metadata we do not correct them. Column names are modified to include the year of data capture.
Solomon and Björk non-APC data	Article counts and subject analysis.

Limitations: open access journals by publishers that were not listed in DOAJ in either 2014 or 2016 are not included. Non-English journals, new journals and journals by new publishers are probably under-represented in this sample. Finding errors (e.g., changes in URL, internet connectivity issues) may have a minor impact on data.

3. Explanation of Specific Data

- WC = Walt Crawford (throughout)

3.1. Journal and Publisher Identification, 2016 Article Processing Charge (APC) and Article Page Processing Charge (APPC) Data, and Associated Currency and Provenance Information

- A to C: journal and publisher identification.
- A and B: Publisher name and Journal title. This data comes from DOAJ 2016 metadata where available, otherwise the publisher's website and occasionally from WC's dataset (journals that were in DOAJ at the time of WC's sampling at the end of December 2015 but had been removed from DOAJ prior to our sampling at the beginning of February 2016).
- C: Journal URL. This data is strictly copied from DOAJ. Any errors are copied from DOAJ and the SKC team has not added URLs for journals not listed in DOAJ.
- D–H: 2016 Article Processing Charge (APC) and I–M Article Page Processing Charge (APPC).
- Columns D to M are the principal focus and contribution of the 2016 update and expansion of the OA APC dataset.
- Column D, "2016 APC publisher website original currency" provides the APC from the publisher's website in the original currency in addition to qualitative information where this information is not appropriate for a given title. This data is derived from lookup on the publisher's website either by the SKC team, WC, or DOAJ. The SKC team gathers data directly from publisher websites for: (a) all journals sampled in 2014 and 2015; (b) additional titles on websites of specific publishers studied; and (c) titles from WC's study where WC provides an amount and there is reason to question whether the original currency is USD (WC's dataset lists prices in USD exclusively) (approximately 1400 titles), e.g., titles published in countries other than the US. The provenance of data not derived by the SKC team is provided in column N.
- The APC amount is determined by the SKC team as follows:
 - costs associated with the publication that have nothing to do with paying for the service of coordinating peer review and publishing the article per se are not included, e.g., subscription rates, offprint sales
 - where differential fees are provided for multiple article types, the type "research article" or closest equivalent is selected
 - where differential fees are provided based on different variations the team aims to select the most common APC for local authors, i.e.,
 - regular fee not discounted fee for society members, low/mid income countries
 - regular fee not including "extras" such as extra pages, editing services, taxes etc.
 - fee for local authors not the fees given for the rest of the world
- Where different currencies are provided, it can be difficult to select a "main" currency; the following criteria are used to select the original currency
 - currency of country of publication
 - first currency listed
 - currency listed in DOAJ (column BJ)
 - currency previously recorded in OA APC project (to facilitate longitudinal comparison)
- The following table explains in detail the data used in columns D and I in alphabetical order (Table 2).

Table 2. Data used in column D (APC) and column I (APPC).

2016 APC Publisher Website Original Currency Qualitative Element	Meaning	Notes	Frequency
active 2015 title not found 2016	this is an expansion of the “title not found” category	added to 2016 dataset to facilitate longitudinal checking, i.e., if titles are not found they will be retained for a to-be-determined length of time in case they are found later on (e.g., change in publisher, temporary website issue)	9
APPC	Article Page Processing Charge i.e., per-page rather than per-article charge. See columns I–M for data	APPC is noted here to ensure data is entered for all journals included in the sample	137
cannot confirm due to language, cannot read (Chinese only)	obvious		2
ceased publication	used when it is clear that the journal is no longer published; distinguished from “title not found” where situation is unclear		3
coming in 2017	title found on publisher's website, forthcoming journal no cost given yet—a form of “cost not specified”		1
conflicting information	used where complexity of model makes it impossible to decide on an APC using the criteria listed above (e.g., price varies with number of authors)		1
cost not specified	Publisher’s website discusses charges but no amount is given		91
no cost found	we were unable to find either an APC or APPC or information indicating that there is none		404
discontinued	see ceased publication		2
F (WC)	no publication fee, data from WC study	temporary distinction from “no publication fee” to highlight question of provenance. This distinction will be removed in the next update	7040
no publication fee	no publication fee, data from SKC team		745
numeric data	numeric data indicates that there is an APC and an amount is given. This includes journals with 0 APC, i.e., the publisher uses the APC model but there currently is no charge	An APC with a number amount is obvious. 0 (zero, number) APC is used to distinguish journals that use the APC model that do not currently charge from journals that have confirmed they do not use the APC model. For example, new journals from publishers like Hindawi and MDPI that are APC-based often have no charges in the first year or two of publication	3282
predecessor	journal has changed title	used to make connection between changed titles for purposes of comparing APCs	1
see next phase of journal	used for Copernicus “discussion” titles, a type of preprint journal; APC is applied in the final (non-discussion) journal		6
title merged with...	indication of succession of journal		1
title not found	could not find the journal, cannot confirm if the journal is ceased or whether this is a finding error, e.g., temporary internet connectivity issue	tried using journal URL and DOAJ lookup. Useful as indicator of OA journal attrition	334
Total			12,060

- Columns E–H Currency and currency conversion: original currency for journals with a numeric amount for APC, conversion rate as of 30 June, and APC converted to USD. Original currency is preferred over USD to facilitate longitudinal comparison, i.e., to avoid conflating price changes and currency fluctuations.

The original currency is determined as follows:

- sole currency when only one currency is given
- local currency (country of publication) where given
- currency selected in previous samples
- first listed currency where multiple currencies are given and local currency cannot be determined
- Conversion rate: determined using the Bank of Canada’s 10-year currency converter with the date of 30 June 2016 where possible and XE Currency Converter for currencies not covered by the Bank of Canada (BDT, EGP, IQD, IRR, NGN, YER).
- Columns I to M provide Article Page Processing Charge (APPC) data details for journals identified as APPC in column D, original currency and currency conversion as described above for columns D to H.
- Column N: Provenance of 2016 APC data if not SKC team lookup. This is a new column designed to facilitate in-depth understanding of the data in column D as this data now comes from 3 sources that use different methods (SKC team, WC, DOAJ). USE WC identifies journals where data is from WC’s dataset. DOAJ is for journals where DOAJ APC metadata is used.
- Column O. Publisher type 2015 (historical) contains data obtained during the 2015 data collection period on publisher type plus occasional updates from 2016, i.e., no systematic attempt was made to complete this data, however when information such as society or university sponsorship or commercial partnerships was available, this data was included. See also column BG, society or institution data from DOAJ 2016. See Table 3 below for codes.

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

Column O

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 3 and 4).

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

Column (2015)	Column (2014)	Column (2013)	Column Title	Notes
P-U	AA-V	AB-AG	APC	See above for details.
			APC Original Currency	See APC above for details.
			Rate	Currency exchange rate (to 2 decimals) as of 15 May 2015.
			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 15 May of the year in question.
			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.
			APPC original currency	See APC and APPC below.
			APPC USD	See APC and APPC below.
			APPC USD Value	Copy of column AC with formula removed.
DY	EV		Max. pages per article	Maximum pages per article where indicated. This may be an absolute maximum or a maximum included under the default APC.
DZ	EW		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 is on the publisher's website. NM for not mentioned means that there is no indication as to whether variations in pricing might apply or not.
EA	EX		Premium price for fast track	Optional additional charge to speed up publication
EB	EY		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.
EC	EZ		Language editing	Extra charge if copyediting needed. Sometimes optional, i.e., authors can make their own arrangements or pay for this service through the journal.
ED	FA		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.
EE	FB		Waivers/discounts based on income	Indication that a waiver may be considered in case of hardship (other than medium and low income countries).
EF	FC		Differential pricing for local authors	Whenever different pricing is given for authors in a particular region. In some cases "local" is assumed.
EG	FD		Waivers/discounts for low/medium income countries	A common discount, often based on World Bank country classifications.
EH	FE		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, etc.

Table 4. Cont.

Column (2015)	Column (2014)	Column (2013)	Column Title	Notes
EI	FF		Waivers/discouts based on individual membership in society or association	This is for individual society members as distinct from institutional memberships.
EJ	FG		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.
EK	FH		Submission fee	Fee charged on submission rather than on publication.
EL	FI		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.
EM	FJ		Extra charge for CC-BY (or varies by license type)	Fee varies based on chosen license.
EN	FK		Temporary discounts	Special time-limited offer.
EO	FL		Differential pricing by article type	Pricing varies by article type. For example, a research article may cost more than a case study.
EP	FM		APC only if there is an author fund	APC applies only if the author has access to funding, otherwise free.
EQ	FN		Discounts for Students	Obvious
ER	FO		Discounts for high quality or extra charges for poor quality	Language is typically vague.
ES	FP		Using publisher's template	Lower price if publisher's template is used.
ET	FQ		Different price format Latex/Word/PDF	Pricing is higher or lower depending on format.
EU			Loyalty discount	Repeat author discount.

3.2. Historical APC and APPC Data (2010, 2013, 2014, 2015)

- Columns P–AJ contain historical APC and APPC data. Columns P–U (2015); V–AA (2014) and AB–AG (2013) contain data for the years indicated gathered by the SKC team; see corresponding codes above for columns D–M to interpret.
- Columns AH–AJ contain historical data from 2010 contributed to the project by Solomon and Björk (2012). Note that Solomon and Björk's methods are slightly different from the SKC project. Their study was a random sample limited to APC charging journals in DOAJ at the time of their study. Their sample was random in nature and does not necessarily reflect all the titles by any given publisher that would have had APCs at that point in time. Solomon and Björk estimated per-article costs for journals using the APPC method based on the average number of pages rather than treating as a different model.

3.3. Walt Crawford's 2015 Data

- Columns AK to BC are WC's data from Crawford's [3] Gold Open Access Journals 2011–2015. See citation for detailed description of data. In addition to data added to column D (APC data,

particularly free journals), WC's data includes counts of articles published in each journal sampled per year, country of publication, and potentially useful notes.

3.4. DOAJ Metadata as of 1 February 2016 and 15 May 2014

- Columns BD–DF are metadata downloaded from the DOAJ website as of 1 February 2016, with added information in the column headings to identify the columns as DOAJ columns and indicate the year. Note that DOAJ Journal Title information is found in column B, which also contains title information from publisher websites. To identify the source of data in column B, refer to column BC “Publisher DOAJ 2016 02”. If there is data in this column, the journal title in column B is from DOAJ. If there is no data in this column, the journal title comes from the publisher's website.
- Columns DG–DX are DOAJ metadata from 14 May 2014.

3.5. Historical Qualitative Data

- Columns DY–EU (2015), and EV–FQ (2014) provide qualitative information for journals sampled in this time frame as described below. See also APC and APPC information for these years in columns E–M.

3.6. 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 AA (2014)

AB to AG (2013)

P to U (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.

- Columns FR to FU contain a unique subject analysis developed by the SKC team based on DOAJ metadata from 2014, gathering DOAJ detailed subject headings into larger groupings for purpose of analysis.
- Columns FV to GA are from the study of Solomon and Björk (2012). See also columns AH–AJ.
- Columns GB–GF provide data related to provenance of data in this study, i.e., journals in DOAJ in one of the years sampled but not another, journals with data derived from the publisher's website that were not in DOAJ or in the longitudinal study.

3.7. DG to DW: DOAJ 2014 Metadata

Columns DF to DV are from DOAJ 2014 metadata (at the time of the first annual survey), except for titles added in 2015 that are on the publisher's website but not in DOAJ. Some 2015 data were added. See, for example, column GB “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 GD “not in DOAJ 2014 or 2015”. Titles that only have information in columns DF and DG (publisher and title) are another indication that the title was on the publisher's website but not in DOAJ (Table 5).

Table 5. Columns DG to DX (2014).

Column	Column Title and Notes/Deviations from DOAJ Metadata
DG	Publisher (from DOAJ; occasional clean-up to facilitate gathering, e.g., typo correction).
DH	Title (almost always directly from DOAJ; occasional small variations due to title name or publisher changes).
DI	Title alternative (from DOAJ; 2015 only. These data are not used in the study. Data from 2014 were deleted).
DJ	Identifier. Journal URL.
DK	Language.
DL	ISSN
DM	E-ISSN
DN	Keyword DOAJ 2015
DO	Start year
DP	End year (blank in DOAJ; titles no longer active are removed).
DQ	Added on date (date added to DOAJ or record updated).
DR	Subjects
DS	Country
DT	Publication fee. Blank in DOAJ as of 2014 and 2015. Earlier DOAJ metadata indicate Yes, No, or Conditional here.
DU	Further information. URL for further information on publication fee. Blank in DOAJ as of 2014 and 2015.
DV	CC license
DW	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.
DX	Publication charges DOAJ 2014. Limited data.

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, 2015 data provided by Walt Crawford, 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.; Brutus, W.; Dumais-DesRosiers, M.; Kakou, T.L.; Laprade, K.; Merhi, S.; Salhab, J.; Volkanova, V. & Wheatley, S. Open access article processing charges longitudinal study 2016 dataset [<http://dx.doi.org/10.5683/SP/KC2NBV>]. 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. When posting your own dataset, please include at minimum the journal title and ISSN as these are key matching points for merging together different datasets.

Acknowledgments: The authors gratefully acknowledge funding provided by Canada’s Social Sciences and Humanities Research Council under an Insight Development Grant for 2014–2016 and an Insight Grant for 2016–2021 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. Writing organization and quality control assistant: Salah Merhi. Data gathering and analysis: Widlyne Brutus, Myriam Dumais-DesRosiers, Tanoh Laurent Kakou, Katherine Laprade, Arbia Ouerghi, Jihane Salhab, Victoria Volkanova and Sara Wheatley.

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.; Mondésir, G.; Salhab, J.; Villamizar, C.; Calvé-Genest, A.; Desautels, L. Open Access Article Processing Charges (OA APC) Longitudinal Study 2015 Preliminary Dataset. *Data* **2016**, *1*, 4. [[CrossRef](#)]
2. Morrison, H.; Salhab, J.; Mondésir, G.; Calvé-Genest, A.; Villamizar, C.; Desautels, L. Open Access Article Processing Charges Longitudinal Study 2015. Available online: <https://dataverse.scholarsportal.info/dataverse/oaapc> (accessed on 4 April 2017).
3. Crawford, W. Gold Open Access Journals 2011–2016. 2016. Available online: <http://walt.lishost.org/2016/05/gold-open-access-journals-2011-2015-its-here/> (accessed on 1 April 2017).
4. 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](#)]
5. Morrison, H.; Salhab, J.; Calvé-Genest, A.; Horava, T. Open Access Article Processing Charges: DOAJ Survey May 2014. *Publications* **2015**, *3*, 1–16. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution - NonCommercial - NoDerivatives (CC BY-NC-ND) license (<http://creativecommons.org/licenses/by/4.0/>).