Plan S Price Transparency Framework: Implementation Guide for Publishers

Alicia Wise and Lorraine Estelle, Information Power

Introduction

This transparency framework evolved from a project involving funders, libraries, publishers, and universities to inform the development of implementation guidelines for one of the Plan S requirements: for OA services and prices to be communicated in a transparent, practical, and insightful way. In December 2019 an independent report and draft framework were published. The draft framework was piloted in Q1 2020 by publishers Annual Reviews, Brill, The Company of Biologists, EMBO Press, European Respiratory Society, F1000 Research, Hindawi, IOP Publishing, PLOS, and Springer Nature from whom we learned an enormous amount. We worked with Owen Stephens, a freelance metadata expert, who helpfully refined the framework to help support the automation of data sharing.

Why a transparency requirement?

The first iteration of Plan S published by cOAlition S in 2018 had a focus on setting a maximum price that the funders were prepared to pay publishers for open access publishing services. In the second iteration published in 2019, cOAlition S decided not to standardise and cap the reimbursement of these services. cOAlition S has signalled a willingness to financially support publication fees that are fair and reasonable. More information is needed from publishers for customers to reach a view about what is fair and reasonable.

Providing this information is an opportunity for publishers to repair a breakdown in trust, be responsive to customers, to demonstrate their commitment to open business models and business cultures, to build awareness of their services and value, and to qualify for OA service funding under Plan S.

For customers, funders, libraries and researchers, using the information is an opportunity to better understand if the fees they pay are commensurate with the open publication services delivered, and to better understand the elements of those services, for example peer review, editorial work, copy editing. Institutions will also be able to use these data to inform discussions with researchers, for example to offer advice to them about their publishing options.

Having the information in a central standardised way will help customers enormously. Much of this information is already available to them, but this approach will lower their costs of accessing and comparing the information and will therefore aid competition.

The transparency framework will enable conversations and comparisons that will build confidence amongst customers that prices are fair and reasonable. Ultimately, it will enable publishers to communicate the price of services in a way that is transparent, practical to implement, and insightful.

How will this work in practice?

This is an evolving framework subject to periodic review and open to influence by publishers, service provides, and users. It is not intended as a cost accounting exercise, nor will it be audited.

Publishers are asked to provide these data annually beginning 1 July 2022. Early adopters are welcome and cOAlition S will provide details on where publishers should deposit data by 1 October 2021. Publishers who wish to provide data for 2019 and/or 2020 will be welcome to do so even though this is not required.

We appreciate that price transparency requirements are new, and it will take time for systems to evolve to capture and report on this information. During our pilot we asked participating publishers to let us know where system developments were required so that we could inform platform vendors. It is possible that your manuscript submission system vendor is already aware of these new requirements and considering how best to support customers to implement price transparency requirements. We encourage you to contact them.

Structure and data entry conventions

Data should be provided for each title in a publisher's portfolio. Please create a new row for each journal in your portfolio.

If you have a range of titles that are very similar to one another, and all the data provided for these would be within ±5%, then it is acceptable to average across these titles. In this case you would enter each journal in its own row and provide the same data for each title in the group. You would then complete Column F to indicate which rows have been grouped together to calculate the average figures reported.

The data collection spreadsheet is structured in three sections:

- **Basic title metadata** in columns B to U, shaded blue.
- **Contextual metadata** in columns V to AF, shaded orange. These columns provide you with an opportunity to provide both qualitative and quantitative context about your journals and their nature and quality.
- **Price breakdown information** in columns AG to AN, shaded pink.

In order to ensure that data are collected consistently, and in a way that will be easy to analyse and potentially machine process in the future, the following principles and conventions have been followed when creating the Transparency Framework Data Collection Spreadsheet.

- Each data set (referring to a single publication) should be a single row
- Each field in the document should be specific and record a single type of information
- Field names contain only alphanumeric characters, underscores and hyphens
- Words in field names are separated by hyphens

- Underscores in field names are used to give some degree of hierarchical structure to column headings (e.g. by using underscores in "counter5_unique-item-requests" and "counter5_total-item-requests" it can easily be understood that both these data elements come from the COUNTER Release 5 Guidelines)
- Existing standards should be used for common data types where appropriate, but we also describe how the data can be calculated without reference to these standards.
- Where identifiers for publications, organisations or other entities are required, the identifier scheme used should always be explicit (either in the template or in the response)
- Any measurements (e.g. time periods) should have explicit units of measurement (either in the template or in the response)

Tips for completing the Data Collection Spreadsheet

If you have journals that are very similar to one another, and all the data provided for these would be within + or – 5%, then it is acceptable to average across these. Indicate which rows were grouped and averaged in column F.

Columns AG through AN ask you to indicate what percentage of the prices relate to different elements of value delivered. This is not meant to be a detailed cost accounting exercise, estimates are fine. We trust publishers will make good efforts to provide good estimates. There are opportunities to build understanding and trust, or to lose them if estimates are poor or misleading.

For society-owned journals published by a partner both organisations are likely to need to require information for the Data Collection Spreadsheet to be completed. Be sure to factor in the percentage of price for each organisation in columns AG through AN. For example, societies might receive payments for their development of the journal (column AG), their contributions to peer review (column AI), etc. Any one-off costs such as signing bonuses should be amortised over the lifetime of the agreement and the relevant portion for the reporting period included in the relevant column.

Overheads and profit/surplus should be distributed evenly across columns AG through AN.

For ease of data entry, you might find it helpful to add a column to the spreadsheet so that you can add the title or internal identifier for each journal. There is no need to provide this when you share the data.

Column-by-column guidance

Basic title metadata

Column B issn_publication-format_print

Only one of columns B, C, D, or E is required to identify the title for which transparency data is supplied.

Column C issn_publication-format_electronic

Only one of columns B, C, D, or E is required to identify the title for which transparency data is supplied.

Column D issn-l

Only one of columns B, C, D, or E is required to identify the title for which transparency data is supplied.

Column E journal-id_doi

Only one of columns B, C, D, or E is required to identify the title for which transparency data is supplied.

Column F clustered_titles

To make the framework easier for publishers to implement, it is possible to group titles that have less than + or – 5% variance in all data fields and to report averages. This field is optional and is used to report which rows of data report such averages.

Column G discipline-area

This is a required field. Indicate the number from https://read.oecd-ilibrary.org/science-and-technology/frascati-manual-2015_9789264239012-en#page60 that best describes your journal **or** choose from one of the following: all disciplines, all STEM disciplines, or all HSS disciplines.

Column H owner_ROR-ID

This is an optional field. Where the publication is owned by an organisation other than the publisher (e.g. a Learned Society) then the owner's ROR ID should be supplied.

Column I publisher_ROR-ID

This is a required field. ROR ID of the organisation which is responsible for publishing the title.

Column J apc-list-price_range-low

This is a required field. The lower price of any APC list price range for primary research articles. If there is just a single APC list price enter the same number in both column J and column K.

Column K apc-list-price_range-high

This is a required field. The higher price of any APC list price range for primary research articles. If there is just a single APC list price enter the same number in both column J and column K.

Column L apc-list-price_currency

This is a required field. Indicate what currency your list prices are set in using a valid ISO code. These can be found online at https://en.wikipedia.org/wiki/ISO_4217#Active_codes.

Column M apc-waiver-discount_policy

This is a required field. A link to the publisher's policy for APC waivers and discounts, if there is one.

Column N subscription-list-price_range-low
This is a required field. The lower price of any subscription list price range for primary research articles. If there is just a single subscription list price enter the same number in both column N and column O.

Column O subscription-list-price_range-high
This is a required field. The higher price of any subscription list price range for primary research articles. If there is just a single subscription list price enter the same number in both column N and column O.

Column P subscription-list-price_currency
This is a required field. Indicate what currency your list prices are set in using a valid ISO code. These can be found online at https://en.wikipedia.org/wiki/ISO_4217#Active_codes.

Column Q subscription-list-price_member-of-the-Association-of-American-Universities
This is a required field. To facilitate comparison of list prices between publishers, please tell us the subscription list price for a typical research-intensive university in the US.

Column R subscription-discount-policy
This is a required field. Provide a link to the publisher's policy for discounting subscription prices, if there is one.

NB: If you use forms of pricing other than APCs and subscriptions for your Open Access publishing and platform services, please append additional columns at the end of the template (i.e. starting from column AP) to provide list price and currency information.

Column S date-information-submitted
This is a required field. The date the data in this framework were prepared, using format yyyy-mm-dd.

Column T reporting-period-start
This is a required field. Ideally the data in this framework should report on articles published in the last full calendar year. We recognise that this may not always be possible (e.g. for a title that is less than 1 year old) and so request that you tell us the start date of the period you are reporting using format yyyy-mm-dd.

Column U reporting-period-end

This is a required field. Ideally the data in this framework should report on articles published in the last full calendar year. We recognise that this may not always be possible (e.g. for a title that is less than 1 year old) and so request that you tell us the end date of the period you are reporting using format yyyy-mm-dd.

Contextual metadata

Column V price-transparency-context

This is an optional field. Provide a link to any additional context you would like to make available. This is an opportunity for publishers to share how they have compiled the data for this framework, provide additional insight into how to interpret the figures, or explain why the figures are the way they are. If your journal price does not cover its costs, for example because it is cross-subsidised or benefits from significant revenue streams that are not related to APCs or subscriptions, then this is the place to provide this context if you wish to do so.

Column W research-articles-published
This is a required field. Number of research articles published in the reporting period. Exclude author corrections, corrigendum, errata, editorials, and reviews.

Column X acceptance-rate
This is a required field. The acceptance rate for articles published in the reporting period.

Column Y desk-rejection-rate

This is a required field. Desk rejection rate for the reporting period.

Column Z issue-publication-frequency
This is a required field. Choose the best of the following choices to describe the frequency of publication issue in your journal:

Annual (=one per year)
Bimonthly (=one every two months)
Semiweekly (=two per week)
Daily (=one per day)
Biweekly (=one ever two weeks)
Semiannual (=two per year)
Biennial (=one every two years)
Triennial (=one every three years)
Three times a week
Three times a month
Monthly
Quarterly
Semimonthly (=two per month)
Three times a year
Weekly
Continuously updated
Other

Column AA median-number-reviews
This is a required field. Median number of reviews for research articles published in the reporting period. If an article goes back to the same reviewer for a second review, then this counts as two reviews. If reviews cascaded to the journal with a paper, then count this as zero.

Column AB median-time-submission-to-first-decision

This is a required field. Median time in days from submission to first decision in reporting period. Decision means whether to reject or send for peer review. Include decisions based on assessment by in-house prepublication teams, editorial boards, editorial boards and external referees, subeditors, etc.

Column AC median-time-peer-review
This is a required field. Median time in days that the submitted article is under review using the NISO definition of article versions (https://www.niso.org/publications/niso-rp-8-2008-jav) in reporting period specified by reporting-period-start and report-period-end.

Column AD median-time-acceptance-to-publication
This is a required field. Median time in days from the time a manuscript is accepted to the time the version of record is available using the NISO definition of article versions (https://www.niso.org/publications/niso-rp-8-2008-jav) in reporting period specified by reporting-period-start and report-period-end.

Column AE counter5_unique-item-requests
This is a required field. Unique_Item_Requests is a metric in the COUNTER Release 5 standard for usage reporting and which is widely required in library agreements with publishers. If a publisher does not use the COUNTER Release 5 standard, then here is an explanation of how you can make the same calculation:

> Total_Item_Requests means all usage requests in all formats (PDF, HTML, etc.)
>
> Unique_Item_Requests means usage of unique content, a journal article for example.
>
> So, if a user views the HTML of an article and also downloads the PDF of the same article, this is counted as:
>
> Total_Item_Requests = 2
>
> Unique _Item_Requests = 1

The data should be aggregated across customers to provide global usage figures.

Column AF counter5_total-item-requests
This is a required field. Total_Item_Requests is a statistic in the COUNTER 5 standard for usage reporting and which is widely required in library agreements with publishers. If a publisher does not use the COUNTER 5 standard, then Total_Item_Requests is the same as 'Total Successful Requests' in COUNTER Release 4. If you use neither version of COUNTER, then here is an explanation of how you can make the same calculation:

> Total_Item_Requests means all usage requests in all formats (PDF, HTML, etc.)
>
> Unique_Item-Requests means usage of unique content, a journal article for example.
>
> So, if a user views the HTML of an article and also downloads the PDF of the same article, this is counted as:

Total_Item_Requests = 2

Unique _Item_Requests = 1

The data should be aggregated across customers to provide global usage figures.

Price breakdown information

Column AG price-breakdown_journal-community-development
This is a required field. Note the total of all the "price-breakdown_" columns (i.e. columns AG to AN) should be 100.
In this column please include activities to establish or develop the journal so that articles can be submitted (i.e. identify a need for the journal, its aims and scope development, investment in funding for the field, editorial board costs, commissioning content, competitor analysis, benchmarking, policy development, and portfolio development).

Column AH price-breakdown_submission-first-decision
This is a required field. Note the total of all the "price-breakdown_" columns (i.e. columns AG to AN) should be 100.
In this column include activities move an article from submission to a first decision about whether to reject or send for peer review (e.g. triaging). For information, this column is separated from the next in order to accommodate the Open Platform publishing model.

Column AI price-breakdown_peer-review
This is a required field. Note the total of all the "price-breakdown_" columns (i.e. columns AG to AN) should be 100.
In this column include all peer review activities including specialist reviews but please do not factor in time or resources donated by editors or reviewers (i.e. recruiting and training peer reviewers and the editors who work with them, peer review management and tracking systems). If the journal operates post-publication peer review, please enter 0% in this column.

Column AJ price-breakdown_services-acceptance-publication
This is a required field. Note the total of all the "price-breakdown_" columns (i.e. columns AG to AN) should be 100.
In this column please include activities from the time an article is accepted to the time it is published (i.e. copyediting, formatting, typesetting, proofreading, assigning and depositing DOIs, XML file conversion (if content is not born XML), tagging, quality assurance checks, integration with abstract and index databases/aggregators/repositories/APC management systems, figure re-lettering or other improvements, dealing with article enhancements such as video abstracts, proofing process and author engagement around this, issue compilation, issue line up, printing, alignment with synchronous articles, vendor management, article pipeline management; addition of ORCIDs and other IDs to support funder/institutional disambiguation, checking references, design).

Column AK price-breakdown_services-post-publication
This is a required field. Note the total of all the "price-breakdown_" columns (i.e. columns AG to AN) should be 100.

In this column please include post-publication activities (i.e. handling ethical queries, provision of usage statistics, long-term preservation and access, reader services, postage, inventory and stock control).

Column AL price-breakdown_platform-development-support
This is a required field. Note the total of all the "price-breakdown_" columns (i.e. columns AG to AN) should be 100.
In this column please include platform development, hosting, and support costs including vendor management. One-off development costs should be amortized over 3 years and figures for the relevant reporting period included.

Column AM price-breakdown_sales-marketing
This is a required field. Note the total of all the "price-breakdown_" columns (i.e. columns AG to AN) should be 100.
Please include all sales and marketing activities (i.e. sales teams, sales administration, legal costs for contracts, negotiations with consortia and libraries, sales agents, invoicing, payments collection, debt recovery, bad debt write-offs, integration with and promotion on social media networks, sponsorship).

Column AN price-breakdown_author-customer-support
This is a required field. Note the total of all the "price-breakdown_" columns (i.e. columns AG to AN) should be 100.
Please include activities related to non-editorial author support and library customer support (e.g. helpdesk, usage/impact/other reports, training, author queries about copyright or CC licenses).

Acknowledgments

We are grateful to:

Project steering group members – Ivy Anderson, Chris Banks, Rachel Bruce, Liam Earney, Sarah Greaves, Hannah Hope, Steven Inchcoombe, Jasmin Lange, Robert Kiley, and Falk Reckling

Pilot participants - Mathias Astell, Rachel Bruce, Sophie Estelle, Georgie Field, Richard Gallagher, Sarah Greaves, Steven Inchcoombe, Paul Johnson, Daniel Keirs, Robert Kiley, Bianca Kramer, Anna Kurek, Rachael Lammey, Jasmin Lange, Michael Markie, Claire Moulton, Niamh O'Connor, Stephanie Paalvast, Bernd Pulverer, Elin Reeves, Johan Rooryck, and Katie Ward

R2R workshop participants – Suzanne Abbott, Roheena Anand, Chris Banks, James Barnett, Laura Cox, Jessica Dennesen, Georgie Field, Jen Goodrich, Daniel Jopling, Andrew Kelly, Arend Küster, Richard Bruce Lamptey, Teresa Lewis, Kamran Naim, Erin Osborne-Marti, Stephanie Paalvast, Amy Price, Claire Rawlinson, Tom Relf, Sneha Rhode, David Ross, Sarah Theissen, Jemima Warren, and Cecilia Heyman Widmark

The following standards, schema and sources were consulted in the process of designing the Transparency Framework Data Collection Spreadsheet:

- JATS
- CrossRef API

- MARC21 (specifically serial publication frequency)
- Counter 5 Guidelines
- NISO RP-17-2013, Institutional Identification: Identifying Organizations in the Information Supply Chain
- ROR IDs
- ISO8601
- ISO4217